BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68, C07H 21/04, 21/02, C12N 9/12 — A1

(11) International Publication Number: **WO 99/05325**

(43) International Publication Date: 4 February 1999 (04.02.99)

(21) International Application Number: PCT/US98/15464

(22) International Filing Date: 24 July 1998 (24.07.98)

(30) Priority Data:
60/053,097 25 July 1997 (25.07.97) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 60/053,097 (CIP)
Filed on 25 July 1997 (25.07.97)

(71) Applicant *(for all designated States except US)*: THE TRUSTEES OF BOSTON UNIVERSITY [US/US]; 147 Bay State Road, Boston, MA 02215 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: GOLDSTEIN, Richard, N. [US/US]; 12 Howland Street, Cambridge, MA 02138 (US).

(74) Agents: CARROLL, Peter, G. et al.; Medlen & Carroll, LLP, Suite 2200, 220 Montgomery Street, San Francisco, CA 94104 (US).

(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: METHODS AND COMPOSITIONS FOR DETERMINING SPECIES OF BACTERIA AND FUNGI

(57) Abstract

Methods and compositions are described for methods and compositions for determining the species for an unknown bacterium (or fungus) in a sample. The approach, which utilizes Ribosomal operon sequences, permits one to identify important bacteria (or fungi) pathogens in a clinical setting.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

METHODS AND COMPOSITIONS FOR DETERMINING SPECIES OF BACTERIA AND FUNGI

This application for patent under 35 U.S.C. 111(a) claims priority to Provisional Application Serial No. 60/053,097 filed July 25, 1997 under 35 U.S.C. 111(b). This invention was made with Government Support under Grant Number DK- RO1-AI37728 awarded by the National Institute of Health. The government may have certain rights in the invention.

FIELD OF THE INVENTION

The present invention relates to the identification of species, and in particular, methods and compositions for distinguishing between bacterial and fungal species and determining the identity of bacterial and fungal pathogens in biological samples.

BACKGROUND

The detection and identification of microorganisms recovered from clinical specimens or environmental sources is an important aspect of clinical microbiology, as this information is important to physicians in making decisions related to methods of treatment. In order that a particular microorganism is identified correctly and consistently, regardless of the source or the laboratory identifying the organism, reproducible systems for identifying microorganisms are critical. As stated by Finegold, "The primary purpose of nomenclature of microorganisms is to permit us to know as exactly as possible what another clinician, microbiologist, epidemiologist, or author is referring to when describing an organism responsible for infection of an individual or outbreak" (S. Finegold, "Introduction to summary of current nomenclature, taxonomy, and classification of various microbial agents," *Clin. Infect. Dis.*, 16:597 [1993]).

Classification, nomenclature, and identification are three separate, but interrelated aspects of taxonomy. Classification is the arranging of organisms into taxonomic groups (*i.e.*, taxa) on the basis of similarities or relationships. A multitude of prokaryotic organisms has been identified, with great diversity in their types, and many more organisms being characterized and classified on a regular basis.

Classification has been used to organize the seemingly chaotic array of individual bacteria into an orderly framework. Through use of a classification framework, a new isolate can be more easily be characterized by comparison with known organisms. The choice of

criteria for placement into groups is currently somewhat arbitrary, although most classifications are based on phylogenetic relationships. An example of the arbitrariness of bacterial classification is reflected in the genetic definition of a "species" as being strains of bacteria that exhibit 70% DNA relatedness, with 5% or less divergence within related sequences (Baron *et al.*, "Classification and identification of bacteria," *in Manual of Clinical Microbiology,* Murray *et al.* (eds.), ASM Press, Washington, D.C., pp. 249-264 [1995]).

Generally, identification of a bacterium is based on its overall morphological and biochemical patterns observed in culture. Indeed, this is the primary technique employed today in clinical laboratories. Of course, this approach is flawed by the fact that diverse organisms can have similar morphologies and/or biochemical requirements. Moreover, numerous organisms associated with disease may not be cultured *in vitro*. Indeed, some do not grow well in traditional *in vivo* culture systems, such as cell cultures or embryonated eggs, nor *in vitro* such as various nutrient agars and broths.

What is needed is a more defined system for speciation, and in particular, speciation of bacteria and fungi. Such an approach should be amenable to automation, permitting the approach to be used routinely in a clinical laboratory.

SUMMARY OF THE INVENTION

The present invention relates to the identification of microbial species, and in particular, methods and compositions for determining the species for an unknown bacterium (or fungus) in a sample. The methods and compositions of the present invention permit distinguishing between bacterial species (or between fungal species) and determining the identity of bacterial (or fungal) pathogens in biological samples. The present invention contemplates a method of speciation that does not require the sequencing of nucleic acid from biological samples. Instead, the method is based on detection of heretofore unknown uniquely conserved portions of ribosomal nucleic acid, such portions being conveniently revealed by restriction digestion of DNA encoding ribosomal nucleic acid, *i.e.* rRNA genes.

In one embodiment of the method of the present invention for speciation, the present invention contemplates analysis of one or more so-called Ribosomal operons ("*rrn*") of a clinical isolate, the operon comprising three genes often arranged in the order 16S-23S-5S for prokaryotes (and 18S-5.8S-25S for eukaryotes), with "spacer" DNA separating each gene (hereinafter represented by: 5'-16S - spacer - 23S - spacer - 5S - 3'). The present invention contemplates that the analysis of at least one of these operons in an unknown bacterial or

fungal species (when evaluated for the "signature band sets" of a particular species, the signature bands and methods for determining signature bands herein described) allows for accurate speciation.

It is not intended that the present invention be limited by the technique by which the operons are analyzed. In one embodiment, primers directed to these sequences can be employed in an amplification reaction (such as PCR). On the other hand, these conserved sequences can conveniently be analyzed with restriction enzymes. Specifically, the present invention contemplates digesting bacterial or fungal DNA with one or more restriction enzymes which will produce a piece of nucleic acid which is within (or bounded by) the 5' and 3' ends of the operon. The resulting digestion product will be conserved for any given species and can serve as a "signature" for that particular species (other species having one or more signature bands of a different size).

Specific embodiments of such a method include (but are not limited to) digestion with one or more restriction enzymes so as to produce any one of the following digestion products:

5'- 16S - spacer - 23S - spacer - 5S - 3',

5'- 16S - spacer - 23S - spacer - 3',

5'- 16S - spacer - 23S - 3',

5'- 16S - spacer - 3',

5'- 16S - 3',

5'- spacer - 23S - spacer - 5S - 3',

5'- 23S - spacer - 5S - 3',

5'- spacer - 5S - 3',

5'- 5S - 3',

5'- 23S - 3'

5'- spacer - 23S - spacer - 3', or

5'- spacer - 23S - 3'

The present invention also contemplates a host of variations on the above digestion products by cleaving in the middle of genes and/or in the middle of spacer regions. However, for the convenience of detecting such digestion products by gel electrophoresis, it is preferred that the digestion product (due to the relatively limited resolution level of gel electrophoresis) be at least 200 bp in size (and more preferably between 400 and 3000 bp in size).

In one embodiment, the present invention contemplates digestion of such DNA with restriction enzymes that cut only once in the DNA encoding 16S ribosomal RNA and only

once in the DNA encoding 23S ribosomal RNA. In a preferred embodiment, the present invention contemplates digestion of bacterial DNA using a single restriction enzyme which cuts only once in the DNA encoding 16S ribosomal RNA and only once in the DNA encoding 23S ribosomal RNA.

In one embodiment, the present invention contemplates a method for bacterial speciation, comprising: i) isolation of bacterial DNA from a sample, said DNA comprising DNA encoding 16S and 23S rRNA; ii) digestion of said isolated DNA with one or more restriction enzymes under conditions such that restriction fragments are produced, said restriction fragments comprising a first digestion product of said DNA encoding 16S and 23S rRNA, said first digestion product comprising at least a portion of said DNA encoding 16S rRNA and at least a portion of said DNA encoding 23S rRNA; iii) separation of said restriction fragments (e.g. by gel electrophoresis), iv) detection of said first digestion product.

In another embodiment, the present invention contemplates a method for bacterial speciation, comprising: i) isolation of bacterial DNA from a sample, said DNA comprising DNA encoding 16S and 23S rRNA; ii) digestion of said isolated DNA with one or more restriction enzymes under conditions such that restriction fragments are produced, said restriction fragments comprising first and second digestion products (e.g. signature bands) of said DNA encoding 16S and 23S rRNA, said first digestion product being larger than said second digestion product, and comprising at least a portion of said DNA encoding 16S rRNA and at least a portion of said DNA encoding 23S rRNA; iii) separation of said restriction fragments (e.g. by gel electrophoresis), iv) detection of said first and second digestion products.

In yet another embodiment, the present invention contemplates a method for bacterial speciation, comprising: a) providing i) a first biological sample comprising bacterial DNA from a known bacterial species, and ii) a second biological sample comprising bacterial DNA from a bacterium whose species is unknown; b) isolating i) a first preparation of bacterial DNA from said first sample and ii) a second preparation of bacterial DNA from said second sample, said DNA of said first and second preparations comprising DNA encoding 16S and 23S rRNA; c) digesting, in any order, i) said first preparation of isolated DNA with one or more restriction enzymes under conditions such that a first preparation of restriction fragments are produced, said first preparation of restriction fragments comprising a first digestion product, said first digestion product comprising at least a portion of said DNA encoding 16S rRNA and at least a portion of said DNA encoding 23S rRNA, and ii) said

second preparation of isolated DNA with one or more restriction enzymes under conditions such that a second preparation of restriction fragments are produced, said second preparation of restriction fragments comprising a second digestion product, said second digestion product comprising at least a portion of said DNA encoding 16S rRNA and at least a portion of said DNA encoding 23S rRNA; d) separating, in any order, i) said restriction fragments (e.g. by gel electrophoresis) from said first preparation, and ii) said restriction fragments (e.g. by gel electrophoresis) from said second preparation; and e) comparing of said first and second digestion products.

It is convenient to isolate bacterial DNA by lysis of bacteria to release DNA. It is also convenient to separate restriction fragments by gel electrophoresis, followed by transfer to a membrane for blotting with an oligonucleotide probe.

It is not intended that the present invention be limited by the nature of the sample. The terms "sample" and "specimen" in the present specification and claims are used in their broadest sense. On the one hand they are meant to include a specimen or culture. On the other hand, they are meant to include both biological and environmental samples. These terms encompasses all types of samples obtained from humans and other animals, including but not limited to, body fluids such as urine, blood, fecal matter, cerebrospinal fluid (CSF), semen, and saliva, cells as well as solid tissue (including both normal and diseased tissue). These terms also refers to swabs and other sampling devices which are commonly used to obtain samples for culture of microorganisms. In addition, fluids such as IV fluids, water supplies and the like are contemplates as samples.

It is also not intended that the invention be limited by the particular purpose for carrying out the biological reactions. The present invention is applicable to medical testing, food testing, agricultural testing and environmental testing. In one medical diagnostic application, it may be desirable to simply detect the presence or absence of specific pathogens (or pathogenic variants) in a clinical sample. In yet another application, it may be desirable to distinguish one species or strain from another.

With regard to distinguishing different species, in one embodiment, the present invention contemplates comparing two samples suspected to be different species. In another embodiment, a species that is suspected to have changed or diverged from the parent species is compared with the parent species. For example, a species or strain of bacteria may develop a different susceptibilities to a drug (e.g. antibiotics) as compared to the parent species; rapid

identification of the specific species or subspecies aids diagnosis and allows initiation of appropriate treatment.

It is not intended that the present invention be limited by the means of detection or the means of comparing first and second digestion products. In one embodiment, said digestion products that are separated by gel electrophoresis are probed with a labeled oligonucleotide in a hybridization reaction.

The present invention can be used with particular success when comparing samples. In one embodiment, the present invention contemplates a method of analyzing nucleic acid in biological samples, comprising: a) providing: i) first and second samples comprising bacterial nucleic acid, ii) a restriction enzyme capable of generating a restriction fragment with (or bounded by) the 5' and 3' ends of a bacterial Ribosomal operon b) treating said nucleic acid of each of said two samples under conditions so as to produce restriction fragments; c) separating said restriction fragments; and d) comparing said restriction fragments from said first and second samples.

It is not intended that the present invention be limited by the number or nature of samples compared. Clinical, food, agricultural, and environmental samples are specifically contemplated within the scope of the present invention.

The present invention contemplates using restriction enzymes wherein the corresponding restriction enzyme recognition sequence exists only once in the 16s and 23s nucleic acid. Alternatively, restriction enzymes can be selected based on the known nucleic acid sequences (*see e.g.* Figures 4 and 7).

DEFINITIONS

To facilitate understanding of the invention, a number of terms are defined below.

"Nucleic acid sequence" and "nucleotide sequence" as used herein refer to an oligonucleotide or polynucleotide, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin which may be single- or double-stranded, and represent the sense or antisense strand.

Prokaryotic ribosomes are constructed from 50S and 30S subunits that join together to form a 70S ribosome. The large subunit comprises a single "23S rRNA" molecule and a "5S rRNA" molecule, while the small subunit comprises a single "16S rRNA" molecule.

As used herein, the terms "complementary" or "complementarity" are used in reference to "polynucleotides" and "oligonucleotides" (which are interchangeable terms that refer to a

sequence of nucleotides) related by the base-pairing rules. For example, the sequence "C-A-G-T," is complementary to the sequence "G-T-C-A."

Complementarity can be "partial" or "total." "Partial" complementarity is where one or more nucleic acid bases is not matched according to the base pairing rules. "Total" or "complete" complementarity between nucleic acids is where each and every nucleic acid base is matched with another base under the base pairing rules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods which depend upon binding between nucleic acids.

Ribosomal RNA molecules are characterized bythe presence of numerous sequences that can form complementary base pairs with sequences located else where in the same molecule. Such interactions cause rRNA molecules to fold into three-dimensional configurations that exhibit localized double-stranded regions.

As used herein, the term "gene" means the deoxyribonucleotide sequences comprising the coding region and including sequences located adjacent to the coding region on both the 5' and 3' ends typically for a distance of about 1-3 kb on either end such that the gene corresponds to the length of the full-length mRNA. The sequences which are located 5' of the coding region and which are present on the mRNA are referred to as 5' non-translated sequences. The sequences which are located 3' or downstream of the coding region and which are present on the mRNA are referred to as 3' non-translated sequences. The term "gene" encompasses both cDNA and genomic forms of a gene.

The chromosomal DNA of prokaryotic cells contains multiple copies of the genes coding for rRNAs. For example, the bacterium *E. coli* contains seven sets of rRNA genes. In the rRNA transcription unit of *E. coli*, the three genes are typically arranged in the order 16S-23S-5S, with "spacer" DNA separating each gene (the spacer DNA separating 23S from 16S typically comprises one or more tRNA genes in addition to unencoded).

The terms "homology" and "homologous" as used herein in reference to nucleotide sequences refer to a degree of complementarity with other nucleotide sequences. There may be partial homology or complete homology (*i.e.*, identity). A nucleotide sequence which is partially complementary, *i.e.*, "substantially homologous," to a nucleic acid sequence is one that at least partially inhibits a completely complementary sequence from hybridizing to a target nucleic acid sequence. The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay

(Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for and inhibit the binding (*i.e.*, the hybridization) of a completely homologous sequence to a target sequence under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (*i.e.*, selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (*e.g.*, less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

Low stringency conditions comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l $NaH_2PO_4 \cdot H_2O$ and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.1% SDS, 5X Denhardt's reagent [50X Denhardt's contains per 500 ml: 5 g Ficoll (Type 400, Pharmacia), 5 g BSA (Fraction V; Sigma)] and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 5X SSPE, 0.1% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

Other equivalent conditions may be employed to comprise low stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, *etc.*) and the concentration of the salts and other components (*e.g.*, the presence or absence of formamide, dextran sulfate, polyethylene glycol), as well as components of the hybridization solution may be varied to generate conditions of low stringency hybridization different from, but equivalent to, the above listed conditions. In addition, conditions which promote hybridization under conditions of high stringency can be used (*e.g.*, increasing the temperature of the hybridization and/or wash steps, the use of formamide in the hybridization solution, *etc.*).

When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe which can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described above.

When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe which can hybridize (*i.e.*, it is the complement of) the single-stranded nucleic acid sequence under conditions of low stringency as described above.

As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids using any process by which a strand of nucleic acid joins with a complementary strand through base pairing to form a hybridization complex. Hybridization and the strength of hybridization (*i.e.*, the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementarity between the nucleic acids, stringency of the conditions involved, the T_m of the formed hybrid, and the G:C ratio within the nucleic acids.

As used herein the term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary G and C bases and between complementary A and T bases; these hydrogen bonds may be further stabilized by base stacking interactions. The two complementary nucleic acid sequences hydrogen bond in an antiparallel configuration. A hybridization complex may be formed in solution (*e.g.*, C_0t or R_0t analysis) or between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized to a solid support [*e.g.*, a nylon membrane or a nitrocellulose filter as employed in Southern and Northern blotting, dot blotting or a glass slide as employed in *in situ* hybridization, including FISH (fluorescent *in situ* hybridization)].

As used herein, the term "T_m" is used in reference to the "melting temperature." The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands (the mid-point). The equation for calculating the T_m of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation: $T_m = 81.5 + 0.41(\% G + C)$, when a nucleic acid is in aqueous solution at 1 M NaCl [*see e.g.*, Anderson and Young, Quantitative Filter Hybridization, *in Nucleic Acid Hybridization* (1985)]. Other references include more sophisticated computations which take structural as well as sequence characteristics into account for the calculation of T_m.

As used herein the term "stringency" is used in reference to the conditions of temperature, ionic strength, and the presence of other compounds such as organic solvents, under which nucleic acid hybridizations are conducted. "Stringency" typically occurs in a range from about T_m-5°C (5°C below the T_m of the probe) to about 20°C to 25°C below T_m. As will be understood by those of skill in the art, a stringent hybridization can be used to identify or detect identical polynucleotide sequences or to identify or detect similar or related polynucleotide sequences.

As used herein, the term "amplifiable nucleic acid" is used in reference to nucleic acids which may be amplified by any amplification method. It is contemplated that "amplifiable nucleic acid" will usually comprise "sample template."

As used herein, the term "sample template" refers to nucleic acid originating from a sample which is analyzed for the presence of a target sequence of interest. In contrast, "background template" is used in reference to nucleic acid other than sample template which may or may not be present in a sample. Background template is most often inadvertent. It may be the result of carryover, or it may be due to the presence of nucleic acid contaminants sought to be purified away from the sample. For example, nucleic acids from organisms other than those to be detected may be present as background in a test sample.

"Amplification" is defined as the production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction technologies well known in the art [Dieffenbach CW and GS Dveksler (1995) *PCR Primer, a Laboratory Manual*, Cold Spring Harbor Press, Plainview NY]. As used herein, the term "polymerase chain reaction" ("PCR") refers to the method of K.B. Mullis U.S. Patent Nos. 4,683,195 and 4,683,202, hereby incorporated by reference, which describe a method for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. The length of the amplified segment of the desired target sequence is determined by the relative positions of two oligonucleotide primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction" (hereinafter "PCR"). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified".

With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by several different methodologies (*e.g.*, hybridization with a labeled probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; incorporation of ^{32}P-labeled deoxynucleotide triphosphates, such as dCTP or dATP, into the amplified segment). In addition to genomic DNA, any oligonucleotide sequence can be amplified with the appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications.

Amplification in PCR requires "PCR reagents" or "PCR materials", which herein are defined as all reagents necessary to carry out amplification except the polymerase, primers

and template. PCR reagents nomally include nucleic acid precursors (dCTP, dTTP etc.) and buffer.

As used herein, the term "primer" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, (*i.e.*, in the presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer and the use of the method.

As used herein, the term "probe" refers to an oligonucleotide (*i.e.*, a sequence of nucleotides), whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification, which is capable of hybridizing to another oligonucleotide of interest. A probe may be single-stranded or double-stranded. Probes are useful in the detection, identification and isolation of particular gene sequences. It is contemplated that any probe used in the present invention will be labelled with any "reporter molecule," so that it is detectable using any detection system, including, but not limited to enzyme (*e.g.*, ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. It is not intended that the present invention be limited to any particular detection system or label.

As used herein, the terms "restriction endonucleases" and "restriction enzymes" refer to bacterial enzymes, each of which cut double-stranded DNA at or near a specific nucleotide sequence. Such enzymes can be used to create Restriction Fragment Length Polymorphisms (RFLPs). RFLPs are in essence, unique fingerprint snapshots of a piece of DNA, be it a whole chromosome (genome) or some part of this, such as the regions of the genome that specifically flank ribosomal operons. All such RFLP fingerprints are indicative of the random mutations in all DNA molecules that inevitably occur over evolutionary time. Because of this, if properly interpreted, evolutionary relatedness of any two genomes can be compared based on the fundamental assumption that all organisms have had a common ancestor. Thus, the

greater the difference in RFLP fingerprint profiles, the greater the degree of evolutionary divergence between them (although there are exceptions). With such an understanding, it then becomes possible, using appropriate algorithms, to covert RFLP profiles of a group of organisms (e.g. bacterial isolates) into a phylogenic (evolutionary) tree.

RFLPs are generated by cutting ("restricting") a DNA molecule with a restriction endonuclease. Many hundreds of such enzymes have been isolated, as naturally made by bacteria. In essence, bacteria use such enzymes as a defensive system, to recognize and then cleave (restrict) any foreign DNA molecules which might enter the bacterial cell (e.g. a viral infection). Each of the many hundreds of different restriction enzymes has been found to cut (i.e. "cleave" or "restrict") DNA at a different sequence of the 4 basic nucleotides (A, T, G, C) that make up all DNA molecules, *e.g.* one enzymes might specifically and only recognize the sequence A-A-T-G-A-C, while another might specifically and only recognize the sequence G-T-A-C-T-A, etc. etc. Dependent on the unique enzyme involved, such recognition sequences vary in length, from as few as 4 nucleotides (e.g. A-T-C-C) to as many as 21 nucleotides (A-T-C-C-A-G-G-A-T-G-A-C-A-A-A-T-C-A-T-C-G). From here, the simplest way to consider the situation is that the larger the recognition sequence, the fewer restriction fragments will result as the larger the recognition site, the lower the probability is that it will repeatedly be found throughout the genomic DNA.

In one embodiment, the present invention utilizes the restriction enzyme called *EcoRI* which has a 6 base pair (nucleotide) recognition site. Thus, given that there exist but 4 nucleotides (A,T,G,C), the probability that this unique 6 base recognition site will occur is 4^6, or once in every 4,096 nucleotides. Given that the *H. influenzae* ("*Hi*") genome (chromosome) is approximately 2 x 10^6 bp (base pairs) in length, digestion of this DNA with EcoRI theoretically should yield 488 fragments. This varies significantly from isolate to isolate of *H. influenzae* because of "random mutations" that inevitably occurs over evolutionary time, some of which either destroy an *EcoRI* sequence cutting site, or create a new one. As such, the overall degree of variation in *EcoRI* RFLP profiles among a series of isolates within a given species such as *H. influenzae*, is indicative of the degree of genetic relatedness of these isolates (although there are exception). Using appropriate algorithms, such RFLP profiles are readily converted to "phylogenetic trees" (see e.g. Figure 3) which are simply a diagrammatic figures indicating the evolutionary divergence of isolates from some theoretically common ancestor.

Once the genomic (chromosomal) DNA of a bacterial isolate has been isolated, it is then digested (cut) with an enzyme such as *EcoRI*. Following the digestion, the resultant individual fragments are separated from one another based on their sizes. This can be done by using agarose gel electrophoresis. In essence, during electrophoresis the smaller molecules (DNA fragments) move faster than larger one and thus the resultant separation is a gradient from the largest to the smallest fragments. These can easily be visualized as bands down the electrophoresis gel, from the top to the bottom with the smallest fragments bottom-most.

Using ribotyping methodology, DNA fragments involving the multiple (*e.g.* 6 for the case of *H. influenzae*, 7 for the case of *E. coli*, etc) ribosomal operons and the immediately flanking DNA sequences (genes) can be distinguished by hybridization of the resultant electrophoresis separated DNA fragments with a radioactively labeled ribosomal operon DNA probe. This then reduces the total number of visualized DNA fragments (predicted above to be approximately 488 restriction fragments) to those only including or immediately flanking the RNA operons, about 14 fragments *in toto* for *H. influenzae*. Nonetheless, because of inevitable random background mutation indicative of evolutionary time, with the exception of very recently evolved clones, every independent isolate of *H. influenzae* will have a variant *EcoRI* ribotype RFLP profile. And the more variant, the more distantly related will be any two isolates so compared. In contrast, rigorous conservation of 16S and 23S rRNA sequences makes possible the unique species-specific RFLPs produced according to the methods and compositions of the present invention.

DNA molecules are said to have "5' ends" and "3' ends" because mononucleotides are reacted to make oligonucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a phosphodiester linkage. Therefore, an end of an oligonucleotide is referred to as the "5' end" if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring. An end of an oligonucleotide is referred to as the "3' end" if its 3' oxygen is not linked to a 5' phosphate of another mononucleotide pentose ring. As used herein, a nucleic acid sequence, even if internal to a larger oligonucleotide, also may be said to have 5' and 3' ends. In either a linear or circular DNA molecule, discrete elements are referred to as being "upstream" or 5' of the "downstream" or 3' elements. This terminology reflects the fact that transcription proceeds in a 5' to 3' fashion along the DNA strand.

As used herein, the term "an oligonucleotide having a nucleotide sequence encoding a gene" means a nucleic acid sequence comprising the coding region of a gene, *i.e.* the nucleic

acid sequence which encodes a gene product. The coding region may be present in either a cDNA, genomic DNA or RNA form. When present in a DNA form, the oligonucleotide may be single-stranded (*i.e.*, the sense strand) or double-stranded.

As used herein, the terms "nucleic acid molecule encoding," "DNA sequence encoding," and "DNA encoding" refer to the order or sequence of deoxyribonucleotides along a strand of deoxyribonucleic acid. The order of these deoxyribonucleotides determines the order of amino acids along the polypeptide (protein) chain. The DNA sequence thus codes for the amino acid sequence.

The term "Southern blot" refers to the analysis of DNA on agarose or acrylamide gels to fractionate the DNA according to size, followed by transfer and immobilization of the DNA from the gel to a solid support, such as nitrocellulose or a nylon membrane. The immobilized DNA is then probed with a labeled oligo-deoxyribonucleotide probe or DNA probe to detect DNA species complementary to the probe used. The DNA may be cleaved with restriction enzymes prior to electrophoresis. Following electrophoresis, the DNA may be partially depurinated and denatured prior to or during transfer to the solid support. Southern blots are a standard tool of molecular biologists [J. Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, NY, pp 9.31-9.58].

The term "Northern blot" as used herein refers to the analysis of RNA by electrophoresis of RNA on agarose gels to fractionate the RNA according to size followed by transfer of the RNA from the gel to a solid support, such as nitrocellulose or a nylon membrane. The immobilized RNA is then probed with a labeled oligo-deoxyribonucleotide probe or DNA probe to detect RNA species complementary to the probe used. Northern blots are a standard tool of molecular biologists [J. Sambrook, J. *et al.* (1989) *supra*, pp 7.39-7.52].

The term "reverse Northern blot" as used herein refers to the analysis of DNA by electrophoresis of DNA on agarose gels to fractionate the DNA on the basis of size followed by transfer of the fractionated DNA from the gel to a solid support, such as nitrocellulose or a nylon membrane. The immobilized DNA is then probed with a labeled oligo-ribonuclotide probe or RNA probe to detect DNA species complementary to the ribo probe used.

The term "isolated" when used in relation to a nucleic acid, as in "an isolated oligonucleotide" refers to a nucleic acid sequence that is identified and separated from at least one contaminant nucleic acid with which it is ordinarily associated in its natural source. Isolated nucleic acid is nucleic acid present in a form or setting that is different from that in

which it is found in nature. In contrast, non-isolated nucleic acids are nucleic acids such as DNA and RNA which are found in the state they exist in nature.

As used herein, the term "purified" or "to purify" refers to the removal of undesired components from a sample.

As used herein, the term "substantially purified" refers to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably 75% free, and most preferably 90% free from other components with which they are naturally associated. An "isolated polynucleotide" is therefore a substantially purified polynucleotide.

The term "sample" as used herein is used in its broadest sense and includes environmental and biological samples. Environmental samples include material from the environment such as soil and water. Biological samples may be animal, including, human, fluid (*e.g.*, blood, plasma and serum), solid (*e.g.*, stool), tissue, liquid foods (*e.g.*, milk), and solid foods (*e.g.*, vegetables).

The term "bacteria" and "bacterium" refer to all prokaryotic organisms, including those within all of the phyla in the Kingdom Procaryotae. It is intended that the term encompass all microorganisms considered to be bacteria including *Mycoplasma, Chlamydia, Actinomyces, Streptomyces,* and *Rickettsia.* All forms of bacteria are included within this definition including cocci, bacilli, spirochetes, spheroplasts, protoplasts, *etc.* Also included within this term are prokaryotic organisms which are gram negative or gram positive. "Gram negative" and "gram positive" refer to staining patterns with the Gram-staining process which is well known in the art [Finegold and Martin, Diagnostic Microbiology, 6th Ed. (1982), CV Mosby St. Louis, pp 13-15]. "Gram positive bacteria" are bacteria which retain the primary dye used in the Gram stain, causing the stained cells to appear dark blue to purple under the microscope. "Gram negative bacteria" do not retain the primary dye used in the Gram stain, but are stained by the counterstain. Thus, gram negative bacteria appear red.

DESCRIPTION OF THE DRAWINGS

Figure 1 schematically shows the 6 Ribosomal operons of the genomically sequenced *H. influenzae* strain Rd.

Figure 2 is an autoradiograph of *EcoRI* RFLPs of *H. influenzae* isolates from diverse sources, including the genomically sequenced strain Rd.

Figure 3 is an *EcoRI* based phylogenic tree of a diverse collection of *H. influenzae* isolates (type "a" through "f", and non-typeable) from variant clinical and environmental sources and geographical locales, showing the signature bands of *H. influenza* with this restriction enzyme.

Figure 4 shows the DNA sequence and restriction map of the *H. influenzae* Rd *rrn*A and Rd *rrn*B operons (16s-spacer-23S-spacer-5S), with restriction sites noted for enzymes cutting 5 times or less (an alphabetical list of restriction enzymes that cut *H. influenzae* Rd *rrn*A and Rd *rrn*B operons 5 times or less is set forth in Table 2), with positions of restriction sites indicated. While the genome of *Hi* Rd contains 6 ribosomal operons, all are identical to the sequences presented here for either *rrn*A or *rrn*B.

Figure 5 shows the 7 ribosomal operons of the genomically sequenced *E. coli* strain MG 1655.

Figure 6 is an autoradiograph of the *EcoRI* RFLPs of *E. coli* isolates from diverse sources, including the genomically sequenced strain MG 1655, showing the signature bands for this species using this restriction enzyme.

Figure 7 shows the DNA sequence and restriction map of the *E. coli* MG 1655 *rrn* ("a" through "h") operons (16S-spacer-23S-spacer-5S), with restriction sites noted for enzymes cutting 5 times or less.

Figure 8 is an autoradiograph of RFLP data for *B. cepacia*, showing signature bands for this species.

DESCRIPTION OF THE INVENTION

The present invention relates to the identification of species, and in particular, methods and compositions for determining the species for an unknown bacterium (or fungus) in a sample. The methods and compositions of the present invention permit distinguishing between bacterial species (or between fungal species) and determining the identity of bacterial (or fungal) pathogens in biological samples. In one embodiment, the present invention contemplates the use of restriction enzymes followed by probing with an oligonucleotide capable of hybridizing to fragments comprising at least a portion of DNA encoding 16S and/or 23S rRNA. In this manner, the present invention applies, in one embodiment, the "discriminatory power" of the methodology of ribotyping to the speciation of microbes for the

first time. The potential use of ribotyping as a method for speciation has been completely overlooked.

To date, ribotyping has been applied for the purpose of examining differences WITHIN a species. Specifically, ribotyping has been employed for the purpose of epidemiological 'typing' within a given species, where 'variability' of the ribotype RFLP profiles of individual isolates, one clinical isolate versus another clinical isolate, was of interest for intra-species discriminatory purposes (for example, to determine whether or not bacterial isolates within a known species were from an epidemic cluster involving a single clone spread among patients). As such, the conserved species-specific signature bands were not recognized as relevant. Instead, the variable bands making up the ribotype profile have been of interest for discriminatory epidemiological purposes and phylogenetic tree building.

The present invention, by contrast, generates a species-conserved set of RFLP bands, unique for each species. While of no interest for intra-species discrimination, these species-conserved sets represent precise markers appropriate for inter-species discriminatory purposes (*i.e.* to determine *per se*, the species of a given, unknown isolate - which is a most needed assay in the clinical microbiology lab of a hospital). Since all bacterial species examined by the inventor display a conserved set of species-specific signature RFLP bands, unique for every species, Ribosomal operon-based discrimination of these unique species specific bands represents the most practical means available for speciation of bacteria (in that the method is less tedious and far more applicable - as compared to sequencing - to the clinical microbiology setting).

It must be stressed that the polymorphisms currently exploited in conventional, epidemiological and phylogenetic ribotyping are polymorphisms that are not directly related to ribosomal operon sequences. Rather, because of the conservation of DNA encoding 16s and 23s rRNA within any species, polymorphisms typically result from variation in closest flanking sequences (that is to say, nucleic acid falling outside of the region defined by: 5'-16S - spacer - 23S - spacer - 5S - 3'). This point can be readily illustrated with the strain Hi Rd, because the complete chromosomal sequence of this strain is known. In this regard, it can be seen from Figures 1 and 2 that it is possible to predict the precise size of the 12 different flank sequences generated by an *EcoRI* digestion (or the fragments generated with any other restriction enzyme for that matter) of the 6 *rrn* operons of strain Rd. With such knowledge of the RFLP profile of the sequenced *Hi* strain Rd, using molecular genetic

methods (such as hybridization), it is possible to precisely analyze any alterations from this prototypic ribotype fingerprint as found among other Hi isolates.

From this example with *Hi*, it should be clear that the polymorphisms generated by the conventional ribotyping technique have nothing directly to do with variability of Ribosomal operon sequences. Rather, these polymorphisms result from variations in the neutral genes that are genetically-linked to (*i.e.* that flank) the multiple ribosomal operons encoded by all bacterial chromosomes.

In constrast to conventional ribotyping, the present invention utilizes the Ribosomal operon sequences which vary less than 3% (and more preferably less than 2%) within a species but vary between species. The description of the invention involves the I) Preparation of Nucleic Acid from Samples; II) Selection of A Restriction Enzyme, III) Design of the Probe, IV) Comparing Biological Samples, and V) Speciation In A Clinical Setting.

I. The Preparation of Nucleic Acid

A. DNA Preparation

The nucleic acid content of cells consists of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). With respect to DNA preparation, a variety of preparation schemes are possible. Typically, the steps involved in purification of nucleic acid from cells include 1) cell lysis; 2) inactivation of cellular nucleases; and 3) separation of the desired nucleic acid from the cellular debris and other nucleic acid. Cell lysis may be achieved through various methods, including enzymatic, detergent or chaotropic agent treatment. Inactivation of cellular nucleases may be achieved by the use of proteases and/or the use of strong denaturing agents. Finally, separation of the desired nucleic acid can be achieved by extraction of the nucleic acid with solvents (e.g. phenol or phenol-chloroform); this method partitions the sample into an aqueous phase (which contains the nucleic acids) and an organic phase (which contains other cellular components, including proteins).

B. RNA Preparation

It is preferred that the present invention utilize DNA and restriction enzymes to analyze bacterial and fungal Ribosomal operon conserved sequences. On the other hand, such conserved sequences may also be examined in the form of 16S, 23S and/or 5S rRNA. For example, such rRNA may be used as template in a PCR reaction with primers (typically DNA primers) capable of amplying such rRNA.

It should be stressed, however, that the preparation of RNA is complicated by the presence of ribonucleases that degrade RNA (*e.g.*, T. Maniatis *et al.*, Molecular Cloning, pp. 188-190, Cold Spring Harbor Laboratory [1982]). Furthermore, the preparation of amplifiable RNA is made difficult by the presence of ribonucleoproteins in association with RNA. (*See,* R. J. Slater, *In*: Techniques in Molecular Biology, J.M. Walker and W. Gaastra, eds., Macmillan, NY, pp. 113-120 [1983]).

II. Selection Of A Restriction Enzyme

As noted above, the present invention contemplates in one embodiment that conserved sequences can conveniently be analyzed with restriction enzymes. Specifically, the present invention contemplates digesting bacterial or fungal DNA with one or more restriction enzymes which will produce a piece of nucleic acid which is within (or bounded by) the 5' and 3' ends of the Ribosomal operon. The resulting digestion product will be conserved for any given species and can serve as a "signature" for that particular species (other species having one or more signature bands of a different size).

A variety of restriction enzymes (and corresponding restriction sites) are contemplated. Given the sequence of the Ribosomal operon for any particular species, restriction enzymes can be selected on the basis of primary structure of the DNA. However, in a preferred embodiment, restriction enzymes are selected based on ultraconserved sequences within the Ribosomal operon; these sequences encode rRNA that takes part in the formation of secondary structures and are known to be more highly conserved because they must fold on themselves (forming secondary structures through Watson/Crick hydrogen bonding). Such sequences encoding rRNA involved in secondary structures are known for some organisms and can readily be determined from the primary structure of the ribosomal DNA for other species using commerically available computer programs.

III. Design of The Probe

In the nucleic acid hybridization step of the method of the present invention, the test DNA is denatured and exposed to denatured DNA of known sequence (i.e. "the probe") from a particular organism. The amount of hybridization between the test DNA and known DNA provides an indication of the degree of relatedness between the test and known organisms. An important drawback to this approach is that hybridization between two single DNA strands can occur even when 15% of the sequences are not complementary. Moreover, to

identify appropriate restriction fragments, one must be able to identify restriction fragments that contain only very short regions (as short as 10 bases) of the 16s, 23s or 5S nucleic acid.

Regardless of these constraints, based on the knowledge of the specific Ribosomal operon DNA sequences of a particular species of bacteria which are recognized by particular restriction endonuclease ("RE"), the present invention contemplates a probe that can be designed to ensure a specific reaction.

The most general ribosomal RNA probe substrate applicable is obtained from purification of bulk ribosomal RNA (16S, 23S and 5S) molecules [*See e.g.* LiPuma *et al., J. Pediatrics* 113:859 (1988)]. A more convenient approach is one using a cloned ribosomal operon which is then digested from the cloning vector, separated by electrophoresis, removed from the electrophoretic gel, and then used as probe substrate [*see e.g.* Arthur et al., Infection & Immunity 58:471 (1990)].

The present invention contemplates a variety of methods for labeling probes, including but not limited to isotopically labeling probes. In one embodiment, nick translation is employed. Briefly, the DNA is lightly "nicked" (single-stranded breaks) with DNAase, and a DNA polymerase which can displace strands at nicks polymerizes DNA using the strand that has not been displaced as template. The nucleoside triphosphates are tagged with isotopes (or other detectable groups) and the polymerase introduces such markers into the nicked DNA.

In another embodiment, the probe is made by random priming. Briefly, the DNA is denatured. Thereafter, small, random oligonucleotides, a labeled substrate, buffers and a DNA polymerase which has no 3'-OH editing function are added. The random oligonucleotides hybridize to places on the DNA and serve as primers for the synthesis of new, labeled DNA.

In yet another embodiment, the probe is end labeled. Briefly, either a kinase attaches a labeled phosphate to the 3'-OH of the DNA or a DNA polymerase with 3' editing function is forced to depolymerize from the 3' end; the resulting single-stranded DNA is used as a template to synthesize labeled DNA.

IV. Comparing Biological Samples

The present invention contemplates, in one embodiment, using electrophoresis to separate RFLP fragments for the comparison of the results between samples. Such an approach can utilize control samples or control fragments to ensure the identification of "signature bands" for a particular species. Moreover, it may be convenient to detect ONLY

the signature bands; this can be done by a variety of methods, including but not limited to the isolating of the signature bands (i.e. free of other restriction fragments). Finally, it may be desirable to automate the analysis.

A. Control Samples

In one embodiment, the present invention contemplates a method wherein a sample of a known bacterial or fungal species is treated in parallel with the test sample(s). In such an approach, the known species is treated with the same restriction enzyme(s) and the resulting fragments are placed in a control lane of the gel, permitting comparison of fragments between the control samples and the test sample(s). Likewise the control may comprise other types or combinations of DNA fragments of known size extracted and prepared for this purpose.

B. Control Fragments

While treating a control sample in parallel is readily done, it may be more convenient to run pre-digested control bands along with the test sample(s). In such a case, the restriction fragments from the pre-digested known sample are simply added to a control lane at the time the test samples have been processed to make them ready for gel electrophoesis.

C. Detecting ONLY The Signature Bands

It may be convenient to detect ONLY the signature bands when comparing samples. This can be done by a variety of methods, including but not limited to the isolating of the signature bands (i.e. free of other restriction fragments). In one embodiment, the present invention contemplates using electrophoresis in combination with a means for sizing the fragment (e.g. HPLC or Mass Spectrometry). In such an approach, restriction enzymes can be utilized that generate the smallest fragment so that this fragment (or fragments) will elute from bottom of the gel prior to the other fragments. The eluted fragment can immediately be examined for size to confirm that the signature band is present or absent in the test sample.

Similarly, the gel for gel electrophoresis can be prepared so as to permit the separation of only fragments in the size range of the signature bands. For example, larger bands capable of hybridizing to the probe would remain at the top of the gel (or be only poorly resolve near the top of the gel).

Also, PCR amplification based on primers including a known restriction site in the conserved region followed by hybridization can be employed.

D. Automation

The present invention contemplates the automation of analysis. In this regard, the present invention specifically contemplates the utilization of the Qualicon (a Dupont subsidiary) "RiboPrinter System" - which is a fast automated apparatus that is (with some modifications, including but not limited to, the provision of marker DNA comprising signature bands) amenable to the automation of some of the above-described methods. In operation, single colonies from 8 unknown microbes are inoculatd directly into a sample carrier into which a "DNA pre pack" is added that contains lysis buffer (enzymes to break open bacteria, along with restriction endonucleases for cutting genomic DNA, along with marker DNA molecules for comparative sizing of RFLP profiles). After initial heat inactivation of colonies, followed by cell lysis and restriction of the DNA, the DNA is then automatically extracted and restriction fragments separated according to size by gel electrophoresis, and then transferred to a hybridization membrane. DNA is then automatically hybridized to a labeled ribosomal operon probe, after which a chemiluminescent agent is introduced. Emission of light from hybridized fragments is captured by digitizing camera and stored as image data. Using proprietary algorithms, a RiboPrint pattern for each sample is extracted from the image data. This pattern can then be compared to other RiboPrint RFLP profiles stored in the system. Such results can be generated every 8 hours, with analysis of the next set of 8 samples begun 2 hours after the first.

The present invention also contemplates a new means for resolving species specific ribosomal RNA bands. This involves hybridization in solution following restriction digestion of the unknown chromosomal DNA sample after which unbound chemiluminescent probe is removed and the sample is electrophoresed. At this point, based on the known rate of migration of DNA fragments of variant size, a chemiluminescent detector is used to detect when hybridized restriction fragments chemilumiescently labeled with the rrn probe elute from the electrophoretic gel. Given the elution rate will be determined by speed of migration, and that migration speed for a fragment of a given size is predictable, the time at which the so chemilumiescently labeled hybridized fragment elutes will indicate its size and thus reveal the signature bands indicative of one species or another.

V. Speciation In A Clinical Setting

The present invention specifically contemplates applying the above-described method to medical diagnostic applications. For example, it may be desirable to simply detect the

presence or absence of specific pathogens (or pathogenic variants) in a clinical sample. In yet another application, it may be disirable to distinguish one species from another. This is a process carried out tens of thousands of times daily in clinical medical microbiology laboratories in hospitals throughout the world, albeit without the benefit of the present invention. Indeed, it is the most common diagnostic analysis (test) carried out in the hospital clinical microbiology laboratory.

Identification of a particular species of microbe causing the infection of a particular patient is needed in order to decide how to treat the infection, *e.g.* what type of antibiotic should be used since different species (e.g. *E. coli* versus *Pseudomonas aeruginosa* versus *Haemophilus influenzae* versus *Burkholderia cepacia*) exhibit different profiles of sensitivity versus resistance to the same antibiotic. Likewise speciation may reveal whether there exists a pathogen expressing tissue-damaging toxins.

Currently, speciation is most typically accomplished in the hospital clinical microbiology lab using a combination of phenotypic assays involving: (i) a series of 10-15 biochemical tests for nutrients required and substrates metabolized or catabolized by microbes); (ii) growth on selective growth media, and (iii) others. At best, results from such tests typically take 12 – 24 hours to obtain and sometimes as long as 5 days (by which time many an infected patient has expired). Such test decipher the species involved with approximately 95% of clinical samples.

The present invention, as noted above, contemplates a non-sequencing approach to speciation. This is because an approach involving sequencing (e.g. purification of DNA, PCR amplification of the 16S gene of the ribosomal operon followed by DNA sequencing) is complex, costly and labor intensive. A sequencing approach is likely to be unsuitable to the hospital setting.

That is not to say, however, that sequencing is altogether inappropriate in all settings. For example, when the 16S genes of different species are compared (*e.g. E. coli* versus *H. influenzae* versus *Neisseria meningitidis* versus *Streptococcus pneumoniae* versus *Staphylococcus aureus*, etc), greater than 10% - 15% differences in the 16S genes are revealed. Given such large differences, it is possible to precisely identify the species of microbe in which the gene was found based on such sequencing of the 16S gene DNA.

EXPERIMENTAL

The following examples serve to illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

In the experimental disclosure which follows, the following abbreviations apply: eq (equivalents); M (Molar); μM (micromolar); N (Normal); mol (moles); mmol (millimoles); μmol (micromoles); nmol (nanomoles); gm (grams); mg (milligrams); μg (micrograms); L (liters); ml (milliliters); μl (microliters); cm (centimeters); mm (millimeters); μm (micrometers); nm (nanometers); °C (degrees Centigrade); Ci (Curies); EDTA (ethylenediamine-tetracetic acid); PAGE (polyacrylamide gel electrophoresis); bp (base pair); CPM (counts per minute).

The present invention is applicable to over 20 other species of bacteria.

To prepare bacterial DNA, cells were pelleted from 5 ml overnight culture, washed with 50:20 mM TE buffer [50 mM Tris (pH 8.0), 20 mM EDTA (pH 8.0)) and re-dissolved in 4 ml 50:2 mM TE buffer (50 mM Tris (pH 8.0), 2 mM EDTA (pH 8.0)]. Cells were first incubated with 50 μl lysozyme solution (20 mg/ml) at 4° C for 30 minutes and then incubated with 50 μl proteinase K (20 mg/ml) and 300 μl 10% SDS at 55° C for 5 hours. 1 ml 10% lauroyl sarcosine (acid free) was added to the cell lysate, and the DNA was purified by equilibrium centrifugation in a caesium chloride-ethidium bromide gradient.

Restriction fragment length polymorphism (RFLP) associated with multicopy ribosomal operons was analysed using an *rrnB* probe. For southern blot analysis, the gel was transferred to a nitrocellulose membrane using a Bio-Rad vacuum blotting apparatus. DNA hybridisation procedure was as follows: After Southern blotting, the membrane was baked at 80°C for 30 minutes, placed in a heat-sealable bag with 10-50 ml prehybridisation buffer, heat-sealed and then incubated at 42°C for 5 minutes. Radio-labelled probe was prepared by adding: 32 μls DNA (DNA sample was a fragment cut from a LMP Agarose gel, and initially boiled for 10 min. before using), 10 μls OLB, 2 μls BSA, 5 μls ^{32}P, 2 μls Klenow. Stock was 0.5 mCu in 50 ml (5 ml = 50 mCu). The mixture was incubated for ~5 hours or overnight, in 37°C H_2O bath. Before adding the probe to the blotted nitrocellulose membrane it was boiled for 10 minutes. Tracking dye was added to the DNA probe before boiling. The labelled probe was added to the membrane using a syringe. The bag was resealed and incubated at 42° C for 4-24 hours on a shaker. The membrane was washed repeatedly but not allowed to dry. Autoradiography was then carried out.

EXAMPLE 1

Conserved, species-specific signature bands: novel genetic markers for inter-species discrimination for *H. influenzae*

Availability of the complete sequence of the chromosome of the *Haemophilus influenzae* ("Hi") strain Rd allowed us to predict *a priori* the resultant *EcoRI* RFLP profile generated from the known 6 *rrn* (ribosomal operon) of this strain. As shown in Figure 1, with EcoRI sites occurring once each, in species-conserved 16S and 23S *rrn* gene sequences of each *rrn* operon, two possible internal fragments (16S-spacer-23S) are generated depending on presence or 1 or 2 tRNA sequences within the spacer region between 16S and 23S genes. These two conserved *EcoRI* fragments (1,503 bp and ~1,748 bp) are found among all *Hi* isolates.

Among the >400, putative typable and "NT" (non-typable, *i.e.* unencapsulated) *Hi* isolates (see Table 1) examined by *EcoRI* ribotyping (Figure 3), all serotype "a" through "e" RFLP profiles and 253 of 311 NTHi (non-typable Hi) RFLP profiles contained both signature bands. 53 NTHi RFLP profiles lacked both signature bands, whereas four lacked the 1748 bp signature band and 1 lacked the 1503 bp signature band. All serotype "f" RFLP profiles lacked both signature bands. These 58 NT and 8 serotype f isolates lacking *EcoRI* ribotype signature bands appear not to be members of the species *H. influenzae* but appear to be a new subspecies or species.

As described above, all 8 serotype "f" isolates plus 55 of 58 NTHi Isolates lacking one or more species specific *EcoRI* signature bands appear clustered together in the Figure 3 dendrogram (the phylogenic tree) as a clearly distinct lineage(s) from all of the other *EcoRI* signature band-containing isolates, both serotype "a" through "e" and NT. The branches in the Figure 3 dendogram are representative of the respective serotypes as follows:

Type "a" is represented by branches 22-26.

Type "b" is represented by branches 29-35.

Type "c" is represented by branches 50-54.

Type "d" is represented by branches 22-28.

Type "e" is represented by branches 58-64.

Type "f" is represented by branches 73-88 (comprising a unique lineage).

Based on methods known in the art, such as multi-locus enzyme electrophoresis (MLEE), this was not revealed in previous phylogenetic analyses of *H. influenzae*.

Preliminary 16S *rrn* gene sequencing has confirmed that putative Hi isolates missing the *EcoRI* ribotype species-specific signature band(s) appear to have been mistyped as Hi by clinical microbiology labs providing these isolates.

EXAMPLE 2

Conserved, species-specific signature bands: novel genetic markers for inter-species discrimination for *E. coli*

An analogous experiment to *the H. influenzae* Example 1 shown above is performed with the species *Escherichia coli.* In this experiment, the computer analysis exemplfied by Example 1 for *H. influenzae,* is utilized for the complete genomic sequence of the *E. coli* isolate MG1655 [Blattner, F., Plunkett III, G., Bloch, C., Perna, N., Burland, V., Riley, M. The complete genome sequence of *Escherichia coli* K-12. Science 277 (5331), 1453-1462 1997]. Roughly 160 independently isolated *E. coli* strains from diverse geographical locales and time periods and sources are analysed (representative data is shown in Figure 6). In this case, the conserved *EcoRI* ribotype RFLP bands indicative of species *E. coli* were resolved to be 2.2 Kb in size. The inventor performed the sequence analyses for all seven (7) ribosomal operons (*rrn*A – *rrn*H) of the *E. coli* strain, looking for appropriately conserved restriction endonuclease sites, preferably one each in 16S and 23S RNA genes. A single site for *EcoR1* was found in the 16S region, and also a single *EcoR1* site was found in the 23S region (Figure 5). Sizes of the signature bands of the ribosomal operons in bp are as follows:

2148 bp (*rrn*A);
2151 bp (*rrn*B);
2064 bp (*rrn*C)
2149 bp (*rrn*D);
2067 bp (*rrn*E);
2143 bp (*rrn*G);
4476 bp (*rrn*H).

Knowing the base pair numbers allowed for *a priori* prediction of the *EcoRI* ribotype RFLP profile of the genomically sequenced *E. coli* isolate MG1655. Also, this allowed for the prediction of the conserved, species specific bands represented by the internal fragments between the 16S and 23S *EcoRI* cut sites (Figure 7).

Both the *E. coli* MG1655 strain and other 168 *E. coli* isolates were then tested to determine the genetic diversity. What was found here is variability in ribotype RFLPs with exception of the two conserved *EcoRI* bands. These two conserved *EcoRI* bands make up the *EcoRI* species specific signature.

Among the 185 putative isolates for this study, some were missing the bands that otherwise always clustered around the 2.2 Kb marker (i.e. the 2,065.5 and 2,148 bp bands). The isolates were re-typed (re-speciated) by the clinical microbiology lab. In every case, those isolates missing the 2 *EcoRI* RFLP bands proved NOT to be *E. coli*.

EXAMPLE 3

Conserved, species-specific signature bands: novel genetic markers for intra-species discrimination for *B. cepacia*

An analogous experiment to Examples 1 and 2 shown above is performed with the species *Burkholderia cepacia*. Only in this case, the conserved *EcoRI* ribotype RFLP bands indicative of species *B. cepacia* were resolved to be 4.2 and 2.6 Kb in size (Figure 8). And, as with *E. coli*, whenever an *EcoRI* ribotype characterized isolate in this *B. cepacia* study was found to be missing these RFLP bands, and subsequently examined by the clinical microbiology lab for speciation, it proved NOT to be in the *B. cepacia* species. One of these mis-typed non-*cepacia* isolates is shown in lane 9 of Figure 8. It can be seen here that this isolate is missing the predictable *B. cepacia* species specific *EcoRI* ribotype bands at 4.2 and 2.6 Kb in size. This isolate proved to be another species, *Xanthomonas maltophilia*.

EXAMPLE 4

Comparison Of Signature Bands

In this example, the specific signature bands were compared across the species tested in Examples 1, 2 and 3 above. When comparing the signature bands for *E. coli* versus *B. cepacia* (see Figures 6 and 8) as well as those for *Hi* verus *E. coli* versus *B. cepacia* (see Figure 3), it is clear that these "signature" bands can be used to distinguish one species from another.

Table 1

Lineage	Strain	Sero-type	Infection	Geographic location	Year
1	N-F433	nt	Ear	Finland	1995
	N-F916	nt	Ear	Finland	1995
	N-F402	nt	Ear	Finland	1995
	N-F432	nt	Ear	Finland	1995
	N-F354	nt	Ear	Finland	1995
	N-F374	nt	Ear	Finland	1995
	N-F375	nt	Ear	Finland	1995
	N-F401	nt	Ear	Finland	1995
	N-F1151	nt	Ear	Finland	1995
	N-F1152	nt	Ear	Finland	1995
	N-F1073	nt	Ear	Finland	1995
	N-F1074	nt	Ear	Finland	1995
	N-F1247	nt	Ear	Finland	1995
	N-F1411	nt	Ear	Finland	1995
	N-F164	nt	Ear	Finland	1994
	N-F233	nt	Ear	Finland	1994
	N-A14	nt	Ear	Cleveland, OH	1986
	N-A1484A	nt	Blood	Connecticut	1980's
	b-EOT81	b	Epiglottis	Ottawa, Ont.	1985-87
	d-ELO127	d	Ear	London, Ont.	1985-87
	ND8-1468	nt			
	ND9-1200	nt	Res	Los Angeles, CA	1995
	N-EA70	nt	Sputum	Montreal, Que	1985-87
	N-ELO147	nt	Sputum	London, Ont.	1985-87
	N-A15	nt	Ear	Cleveland, OH	1986
	N-A32	nt	Ear	Cleveland, OH	1984
	N-ESJ209	nt	CSF	Montreal, Que.	1985-87
	N-F1003	nt	Ear	Finland	1995
	N-F1004	nt	Ear	Finland	1995
	N-F1042	nt	Ear	Finland	1995
	ND21-1328	nt			
	ND23-1109	nt	Res	Hartford, CT	1995
	ND171182	nt	Res	Phoenix, AZ	1995
	ND21038	nt	CSF	Detroit, MI	1995
	ND10-238	nt	Res	Clackamas, OR	1994
	ND11086	nt	Ear	Cleveland, OH	1995
	ND111123	nt	Res	Stanford, CA	1995
	ND17-783	nt			
	ND27-1433	nt			
	ND27-999	nt			
	ND231110	nt	CSF	Hartford, CT	1995
	ND241019	nt	Res	Rochester, NY	1995
	ND271007	nt	Res	Worcester, MA	1994
	ND28-10	nt	Res	New York, NY	1994
	ND28278	nt	Res	New York, NY	1994
	ND7-1526	nt			
2	N-F1363	nt	Ear	Finland	1995
3	ND131158	nt	Ear	Houston, TX	1995

Table 1, cont'd

4	b-B1324	b			1984
	N-EF147	nt	Sputum	Halifax, NS	1985-87
	N-EOT14	nt	Epiglottis	Ottawa, Ont.	
	ND41093	nt	Res	Rochester, NY	1994
5	ND41094	nt	Res	Rochester, NY	1994
6	ND171184	nt	Res	Phoenix, AZ	1995
	ND171186	nt	Res	Phoenix, Az	1995
	ND131157	nt	Res	Houston, TX	1995
	ND171183	nt	Res	Phoenix, AZ	1995
	ND10-239	nt	Res	Clackamas, OR	1994
	ND111124	nt	Res	Stanford, CA	1995
	ND12-1119	nt	CSF	Seattle, WA	1995
	ND13-112	nt	Res	Houston, TX	1994
	N-F567	nt	Ear	Finland	1995
	N-F570	nt	Ear	Finland	1995
	N-F51	nt	Ear	Finland	1994
	N-F54	nt	Ear	Finland	1994
	N-F608	nt	Ear	Finland	1995
	N-F699	nt	Ear	Finland	1995
	N-F732	nt	Ear	Finland	1995
	ND10-189	nt	Res	Clackamas, OR	1994
	b-EA163	b	Sputum	Montreal, Que.	1985-87
	b-EE184	b	CSF	Winnipeg, Man.	1985-87
	ND91197	nt	Res	Los Angeles, CA	1995
	ND3110	nt	Res	Evanston, IL	1994
	ND9-1196	nt	Res	Los Angeles, CA	1995
	ND18174	nt	Res	Chapel, NC	1994
	ND221154	nt	Res	Boston, MA	1995
	ND27-1354	nt			
	ND31031	nt	Res	Evanston, IL	1995
	N-F1015	nt	Ear	Finland	1995
	N-F1158	nt	Ear	Finland	1995
	N-A7	nt	Ear	St. Louis, MO	1985
	N-EF79	nt	Bronchoscopy	Halifax, NS	1985-87
	N-A1276	nt	Blood	St. Louis, MO	1980's
	N-A1328	nt	Blood	St. Louis, MO	1980's
	N-A1636	nt	Blood	St. Louis, MO	1980's
	N-A3247A	nt	Ear	Cleveland, OH	1980's
	N-F1440	nt	Ear	Finland	1995
	N-F187	nt	Ear	Finland	1994
	N-F1159	nt	Ear	Finland	1995
	N-F1382	nt	Ear	Finland	1995
	N-F188	nt	Ear	Finland	1994
	N-F258	nt	Ear	Finland	1994
	N-F261	nt	Ear	Finland	1994
	N-F279	nt	Ear	Finland	1994
7	N-A49	nt	Blood	St. Louis, MO	1995
8	N-A1509	nt	Ear	Philadelphia, PA	1980's
	N-A1512A	nt	Ear	Philadelphia, PA	1980's
9	b-B7109	b	Nasal	Stockholm	1985
10	ND41101	nt	Res	Rochester, NY	1994
11	a-ELO16	a	Eye	London, Ont.	1985-87
	f-EF136	f	Sputum	Halifax, NS	1985-87

Table 1, Cont'd

	ND231115	nt	Res	Hartford, CT	1995
	N-A12	nt	Ear	Cleveland, OH	1985
	ND21041	nt	CSF	Detroit, MI	1995
12	N-F1146	nt	Ear	Finland	1995
	N-F1200	nt	Ear	Finland	1995
	ND9-1201	nt	SA	Los Angeles, CA	1995
	N-F1268	nt	Ear	Finland	1995
	N-F84	nt	Ear	Finland	1994
	ND1-1080	nt	Res	Cleveland, OH	1995
	ND241026	nt	Res	Rochester, NY	1995
13	a-EC195	a	Nasopharynx	Regina, Sask.	1985-87
	N-A1635	nt	Blood	St. Louis, MO	1980's
	ND9-1199	nt	SA	Los Angeles, CA	1995
	ND271002	nt	Res	Worcester, MA	1994
	ND9-1194	nt	Res	Los Angeles, CA	1995
	N-EA145	nt	Sputum	Montreal, Que.	1985-87
	ND21-1206	nt	Res	Mobile, AL	1995
	ND23-938	nt			
	ND231108	nt	Res	Hartford, CT	1995
14	ND271003	nt	BF	Worcester, MA	1995
15	ND271005	nt	Res	Worcester, MA	1994
16	ND211208	nt	Res	Mobile, AL	1995
17	N-A30	nt	Ear	Cleveland, OH	1983
18	a-B6059	a	Sputum	Newcastle, UK	1964
	a-B6064	a	Sputum	Newcastle, UK	1966
	ND9-1195	nt	Res	Los Angeles, CA	1995
	ND41098	nt	Res	Rochester, MN	1994
	ND41100	nt	BF	Rochester, MN	1994
	ND10241	nt	Res	Clackamas, OR	1994
	ND111122	nt	Res	Stanford, CA	1995
	N-F1124	nt	Ear	Finland	1995
	ND10187	nt	Res	Clackamas, OR	1994
	ND271004	nt	CSF	Worcester, MA	1994
	ND28-13	nt	Res	New York, NY	1994
	ND3111	nt	Res	Evanston, IL	1994
	ND41096	nt	Res	Rochester, MN	1994
19	a-B6062	a	Nasal	Newcastle, UK	1965
	c-EC181	c	Sputum	Regina, Sask.	1985-87
	ND271000	nt	Res	Worcester, MA	1994
	N-F979	nt	Ear	Finland	1995
	N-F981	nt	Ear	Finland	1995
	N-A28	nt	Ear	Cleveland, OH	1983
	N-F1071	nt	Ear	Finland	1995
	N-F241	nt	Ear	Finland	1994
	N-F253	nt	Ear	Finland	1994
20	ND21036	nt	CSF	Detroit, MI	1995
21	ND12-1117	nt	CSF	Seattle, WA	1995
	ND3109	nt	Res	Evanston, IL	1994
	N-F639	nt	Ear	Finland	1995
	ND10-240	nt	Res	Clackamas, OR	1994
	d-EOT156	d	Eye	Ottawa, Ont.	1985-87
	N-A1515	nt	Ear	Philadelphia	1980's
	a-EI111	a	Bronchoscopy	Ste.-Foy, Que.	1985-87

Table 1, Cont'd

	c-EC86	c	Ear	Regina, Sask.	1985-87
	N-A26	nt	Ear	Cleveland, OH	1982
	N-A27	nt	Ear	Cleveland, OH	1982
	N-EF105	nt	Sputum	Halifax, NS	1985-87
	N-ESJ136	nt	Ear	Montreal, Que.	1985-87
22	N-F1542	nt	Ear	Montreal, Que.	1985-87
	ND241023	nt	Blood	Rochester, NY	1995
	N-F1396	nt	Ear	Finland	1995
	N-F1541	nt	Ear	Finland	1996
	N-F1233	nt	Ear	Finland	1995
	N-F1241	nt	Ear	Finland	1995
	N-F1275	nt	Ear	Finland	1995
	N-F1345	nt	Ear	Finland	1995
	a-B7190	a	CSF	Malaysia	1973
	d-B6137	d	Throat	Newcastle, UK	1963
	ND271009	nt	Res	Worcester, MA	1994
	N-F1125	nt	Ear	Finland	1995
	N-F1142	nt	Ear	Finland	1995
	N-F1207	nt	Ear	Finland	1995
	N-F1209	nt	Ear	Finland	1995
23	a-B7032	a	CSF	Papua New Guinea	
24	ND171189	nt	Res	Phoenix, AZ	1995
25	a-B6069	a	Throat	Newcastle, UK	1962
	a-B7115	a		Santo Domingo	
26	a-B7421		Nasal	Kenya	
	d-B1168	d		Massachusetts	1983
	a-B1042	a	CSF	Arizona	1981
	a-B7031	a	CSF	Papua New Guinea	
	d-B6150	d	Sputum/blood	Kent, UK	1985
	d-B7033	d	Blood	Papua New Guinea	
	Rd-ATCC51907				
	Rd-RM118				
27	d-ATCC9332				
28	N-ELO79	nt	CSF	London, Ont.	1985-87
	N-OT9	nt	CSF	Ottawa, Ont.	1985-87
29	b-EA122	b	CSF	Montreal, Que.	1985-87
	N-F206	nt	Ear	Finland	1994
30	b-ATCC9795				
31	b-B8069	b	Blood	Knots Landing	1985
32	b-CMINNA	b		Minnasota	
33	ND1-1079	nt	Res	Cleveland, OH	1995
	ND12-898	nt			
	N-F1008	nt	Ear	Finland	1995
	N-F487	nt	Ear	Finland	1995
	ND21040	nt	Ear	Detroit, MI	1995
	N-A9	nt	Ear	St. Louis, MO	1985
	N-F1007	nt	Ear	Finland	1995
34	b-B6094	b	CSF	Wycombe, UK	1985
	b-B7004	b	CSF	Holland	
	b-BEAGAN	b			
	b-B7853	b	CSF	Maryland	1990
	b-B8012	b	Blood	7 Mile Ja	1984

Table 1, Cont'd

	b-B7017	b	CSF	Ghana	1983
	b-B7118	b	Blood	Melbourne, A	1985
	b-B7651	b	CSF	Norway	1980's
	b-B7717	b		Australia	1989
35	N-F430	nt	Ear	Finland	1995
	N-F566	nt	Ear	Finland	1995
	N-F412	nt	Ear	Finland	1995
	N-F413	nt	Ear	Finland	1995
	b-B6107	b	CSF	Oxford, UK	1985
	b-B7020	b	CSF	Ghana	1983
	ND241022	nt	Res	Rochester, NY	1995
	b-B7414	b		Kenya	
	b-EC129	b			
	N-F285	nt	Ear	Finland	1994
	N-F286	nt	Ear	Finland	1994
36	a-B6073	a	Sputum	Newcastle, UK	1966
	a-B6083	a	Sputum	Newcastle, UK	
	c-B6134	c	Abcess	Oxford, UK	1975
37	a-ATCC9006	a			
38	a-B7416	a	Nasal	Kenya	
39	N-A1510	nt	Ear	Philadelphia, PA	1980's
40	ND241028	nt	CSF	Rochester, NY	1995
	ND941	nt	Blood	Los Angeles, CA	1994
	ND111121	nt	BF	Stanford, CA	1995
	ND241025	nt	CSF	Rochester, NY	1995
	N-A16	nt	Ear	Cleveland, OH	1986
	N-A17	nt	Ear	Cleveland, OH	1986
	N-A820A	nt	CSF	St. Louis, MO	1980's
	N-F658	nt	Ear	Finland	1995
41	ND17-1188	nt	Res	Phoenix, AZ	1995
	ND231111	nt	Res	Harford, CT	1995
	ND28279	nt	Res	New York, NY	1994
	N-A1396A	nt	CSF	Minneapolis	1980's
	N-F723	nt	Ear	Finland	1995
42	ND18178	nt	BAL	Chapel, NC	1994
43	ND171185	nt	Res	Phoenix, AZ	1995
	ND28-12	nt	Res	New York, NY	1994
44	N-A24	nt	Ear	Cleveland, OH	1982
	N-A3246A	nt	Ear	Cleveland, OH	1980's
	N-EC194	nt	Ear	Regina, Sask.	1985-87
	ND111125	nt	Res	Stanford, CA	1995
45	d-EF33	d	Sputum	Halifax, NS	1985-87
	N-A1878B	nt	Ear	St. Louis, MO	1980's
	b-EE53	b	Sputum	Nalifax, NS	1985-87
	c-B7424	c		Kenya	
	ND3107	nt	Res	Evanston, IL	1994
	ND3108	nt	Res	Evanston, IL	1994
	ND18177	nt	Res	Chapel, NC	1994
	ND20-144	nt	Res	Decteur, GA	1994
	ND241027	nt	CSF	Rochester, NY	1995
	ND25-209	nt	CSF	Washington, DC	1994
	N-F137	nt	Ear	Finland	1994
	N-EI71	nt	Sputum	Ste.-Foy, Que.	1985-87

Table 1, Cont'd

	N-A5	nt	Ear	St. Louis, MO	1985
	N-EE165	nt	Eye	Winnipeg, Man.	1985-87
	N-F973	nt	Ear	Finland	1995
	ND1-1078	nt	Res	Cleveland, OH	1995
	ND10-188	nt	Res	Clackamas, OR	1994
	ND18176	nt	Res	Chapel, NC	1994
46	ND41095	nt	Res	Rochester, MN	1994
	ND8-1175	nt	Res	St. Louis, MO	1995
47	b-EOT165	b	CSF	Ottawa, Ont.	1985-87
	c-B8032	c			1983
	b-ELO29	b	CSF	London, Ont.	1985-87
	b-ELO38	b	CSF	London, Ont.	1985-87
	N-F1117	nt	Ear	Finland	1995
	N-F486	nt	Ear	Finland	1995
	ND31032	nt	Res	Evanston, IL	1995
	a-B7205	a	CSF	Gambia	1984
	a-EE163	a	CSF	Winnipeg, Man.	1985-87
48	ND221153	nt	Res	Boston, MA	1995
49	ND25-489	nt	CSF	Washington, DC	1994
	ND301076	nt	BF	Syracuse, NY	1995
50	c-EOT36	c	CSF	Ottawa, Ont.	1985-87
	N-A1136B	nt	Blood	St. Louis, MO	1980's
	c-B1271	c		Chicago, IL	1968
	c-B7267	c	Sputum	Malaysia	1973
	N-EOT126	nt	CSF	Ottawa, Ont.	1985-87
	b-ESJ133	b	CSF	Montreal, Que.	1985-87
	c-B1167	c		Massachusetts	
51	c-B7270	c	Sputum	Malaysia	1975
52	ND21-1204	nt	SA	Mobile, AL	1995
53	ND12-599	nt			
54	c-B6132	c	Nasal	Newcastle, UK	1964
	c-ATCC9007	c			
	c-B6129	c		Wellcomb Res. Lab.	1970
55	N-A31	nt	Ear	Cleveland, OH	1983
	N-CBCH-2	nt		Boston, MA	
	a-EC140	a	Ear	Regina Sask.	1985-87
	N-A1514A	nt	Ear	Philadelphia, PA	1980's
	ND25-62	nt	CSF	Washington, DC	1994
	ND3-1553	nt			
	N-F176	nt	Ear	Finland	1994
	N-F477	nt	Ear	Finland	1995
	N-F478	nt	Ear	Finland	1995
	ND13-113	nt	Res	Houston, TX	1994
56	ND2-1037	nt	SA	Detroit, MI	1995
57	ND6-1453	nt			
	ND6-1490	nt			
	ND3-1552	nt			
	ND41097	nt	Res	Rochester, MN	1994
	ND18175	nt	Res	Chapel, NC	1994
	ND18179	nt	Res	Chapel, NC	1994
	ND21039	nt	CSF	Detroit, MI	1995
	ND23-926	nt			
	N-F1104	nt	Ear	Finland	1995

Table 1, Cont'd

	N-F1106	nt	Ear	Finland	1995
	N-EA57	nt	Sputum	Montreal, Que.	1985-87
	N-F1061	nt	Ear	Finland	1995
	c-EC117	c	Tracheal	Regina, Sask.	1985-87
	e-EF142	e	Eye	Halifax, NS	1985-87
	N-CBCH-1	nt		Boston, MA	
	N-CBCH-3	nt	Nasopharynx	Boston, MA	
	N-F1232	nt	Ear	Finland	1995
	N-F758	nt	Ear	Finland	1995
	N-F1147	nt	Ear	Finland	1995
	N-F1231	nt	Ear	Finland	1995
	N-F886	nt	Ear	Finland	1995
	ND16-1529	nt	CSF	Dallas, TX	1995
	ND171187	nt	Res	Phoenix, AZ	1995
	ND18-984	nt			
58	N-EA73	nt	Ear	Montreal, Que.	1985-87
	ND121120	nt	Ear	Seattle, WA	1995
59	ND28-15	nt	BF	New York, NY	1994
60	e-B6181	e	Sputum	Newcastle, UK	1965
61	e-B6168	e		Newcastle, UK	1964
	e-B6169	e	Sputum	Newcastle, UK	1966
	e-B7066	e	Lung asp.	Papua New Guinea	
62	ND231113	nt	Blood	Harford, CT	1995
63	N-F740	nt	Ear	Finland	1995
64	e-B8031	e	Throat	Canyon Bay, USA	1983
	ND21-1203	nt	Ear	Mobile, AL	1995
	e-B7287	e	Sputum	Malaysia	1973
	e-B7423	e	Nasal	Kenya	
	ND221152	nt	BF	Boston, MA	1995
	ND241020	nt	Res	Rochester, NY	1995
	e-ATCC8142	e			
	e-B1018	e		Indiana, USA	1987
	e-B6158	e	Sputum	Newcastle, UK	1962
	e-B6229	e	Sputum	Oxford, UK	1977
65	ND1-1081	nt	Res	Cleveland, OH	1995
66	ND8-102	nt	BF	St. Louis, MO	1994
67	ND12-1116	nt	Ear	Seattle, WA	1995
68	ND41099	nt	Res	Rochester, MN	1994
69	ND301072	nt	Ear	Syracuse, NY	1995
70	ND11083	nt	CSF	Cleveland, OH	1995
71	N-A11	nt	Ear	Cleveland, OH	1985
72	N-A1396B	nt	CSF	Minneapolis	1980's
73	N-A3837B	nt	Ear	Cleveland, OH	1980's
	N-F199	nt	Ear	Finland	1994
	ND271006	nt	Res	Worcester, MA	1994
	N-F200	nt	Ear	Finland	1994
	N-F218	nt	Ear	Finland	1994
74	ND1-1077	nt	Res	Cleveland, OH	1995
75	NL-EOT149	nt	Sputum	Ottawa, Ont.	1985-87
76	ND28-11	nt	Res	New York, NY	1994
77	ND11085	nt	Res	Cleveland, OH	1995
78	N-F1181	nt	Ear	Finland	1995

Table 1, cont'd.

	N-F1251	nt	Ear	Finland	1995
	ND23134	nt	Res	Hartford, CT	1994
	N-F942	nt	Ear	Finland	1995
	N-F943	nt	Ear	Finland	1995
	N-F1292	nt	Ear	Finland	1995
	N-F1306	nt	Ear	Finland	1995
	N-F1414	nt	Ear	Finland	1995
	N-F1543	nt	Ear	Finland	1995
79	N-F1180	nt	Ear	Finland	1995
	ND18171	nt	CSF	Chapel, NC	1994
80	f-B7290	f	Sputum	Malaysia	1974
	ND231114	nt	Ear	Hartford, CT	1995
81	f-B6255	f	Sputum	Newcastle, UK	1967
	f-B7283	f	Sputum	Malaysia	1972
	f-ATCC9833	f			
	f-B6237	f	Nasal	Newcastle, UK	1963
	N-F553	nt	Ear	Finland	1995
	ND18172	nt	Res	Chapel, NC	1994
82	N-EC105	nt	Sputum	Regina, Sask.	1985-87
	f-ELO117	f	Eye	London, Ont.	1985-87
	f-EOT203	f	Eye	Ottawa, Ont.	1985-87
83	N-F161	nt	Ear	Finland	1994
	N-F162	nt	Ear	Finland	1994
84	f-B6252	f	Nasal	Newcastle, UK	1966
85	N-F167	nt	Ear	Finland	1994
	ND31033	nt	Res	Evanston, IL	1995
86	N-A1511	nt	Ear	Philadelphia	1980's
	ND10-242	nt	Res	Clackamas, OR	1994
87	ND271008	nt	Res	Worcester, MA	1994
88	ND241024	nt	Res	Rocester, NY	1995
	ND271001	nt	Res	Worcester, MA	1994
	ND23133	nt	Res	Hartford, CT	1994
	ND241021	nt	Res	Rocester, NY	1995
	b-EOT22	b	Sputum	Ottawa, Ont.	1985-87
	N-F1017	nt	Ear	Finland	1995
	ND31030	nt	Res	Evanston, IL	1995
	N-F599	nt	Ear	Finland	1995
	N-F667	nt	Ear	Finland	1995
	N-F708	nt	Ear	Finland	1995
	ND20-143	nt	Res	Decteur, GA	1994
89	DK-1 E. COLI				
90	ATCC27088 AP				

Table 2

List of restriction enzymes (alphabetical order), cutting *H. influenzae* Rd *rrn*A 5 times or less, with positions of restriction sites indicated.

Enzyme	Freq	Position(s)				
Aat II G ACGT C C TGCA G	1	1180				
Aco III TGCGCA ACGCGT	1	385				
Acc I GT MK AC CA KM TG	3	1241	1586	5077		
Acc III T CCGG A A GGCC T	2	1297	3757			
Ace II G CTAG C C GATC G	1	3992				
Ace III CAGCTCNNNNNNN NNNN GTCGAGNNNNNNN NNNN	3	1074	3788	4825		
Acr I CYCGRG GRGCYC	5	1377	2468	3917	4259	5143
Afa24R I GCCGGC CGGCCG	1	3968				
Aff III A CRYG T T GYRC A	3	679	1223	2644		
Aff IV AGTACT TCATGA	2	652	2718			
Age I A CCGG T T GGCC A	1	4007				
Alw I GGATCNNNN N CCTAGNNNN N	3	1533	1830	4846		
AlwN I CAG NNN CTG GTC NNN GAC	2	1046	4492			
Ama I TCGCGA AGCGCT	1	1346				
Aos III CCGCGG GGCGCC	2	522	4304			
Apo I G GGCC C C CCGG G	2	927	2675			
Aqu I C YCGR G G RGCY C	5	1378	2469	3918	4260	5144
Ase I AT TA AT TA AT TA	1	1890				

Table 2 cont'd

Wednesday, July 22, 1998 2:20 PM
mA RE sites Map (1 > 5519) 484 Cut Sites Site Summary by Enzyme

Enzyme	Freq	Position(s)				
Asp52 I AAGCTT TTCGAA	3	77	2332	4589		
Asp5H I GCATGC CGTACG	1	213				
Asp78 I AGGCCT TCCGGA	1	412				
Ate I CCATGG GGTACC	2	1406	2952			
AtuC I TGATCA ACTAGT	1	11				
Ava I C YCGR G G RGCY C	5	1378	2469	3918	4260	5144
Avr II C CTAG G G GATC C	1	821				
Bae I ACNNNNGTAYC TGNNNNCATRG	2	788	4630			
Bae I NNNNNNNNNNNNNNNACNNN NNNNNNNNNNNNNNNTGNNN	2	814	4656			
Bal I TGG CCA ACC GGT	1	4245				
Ban I G GYRC C C CRYG G	3	847	4755	5508		
Ban II G RGCY C C YCGR G	4	232	927	1008	2675	
BavI CAG CTG GTC GAC	2	1794	4024			
Bbf7811 I TCCGGA AGGCCT	2	1296	3756			
Bbr I A AGCT T T TCGA A	3	78	2333	4600		
Bbs I GAAGACNN NNNN CTTCTGNN NNNN	3	1548	3567	5278		
Bce83 I CTTGAGNNNNNNNNNNNNN GAACTCNNNNNNNNNNNNN	3	3121	3929	5153		
Bcg I GCANNNNNNTCGNNNNNNNN CGTNNNNNNAGCNNNNNNNN	1	1072				

Table 2 cont'd

Wednesday, July 22, 1998 2:20 PM
rmA RE sites Map (1 > 5519) 484 Cut Sites Site Summary by Enzyme

Enzyme	Freq	Position(s)			
Bcg I NN NNNNNNNNNGCANNN NN NNNNNNNNNCGTNNN	1	1038			
Bcl I T GATC A A CTAG T	1	12			
Bco102 II GAAGAC CTTCTG	3	1540	3559	5270	
Bco163 I CTRYAG GAYRTC	2	1621	4338		
Bco35 I CTGGAG GACCTC	2	1167	2984		
BcuI ACTAGT TGATCA	2	3383	3389		
Bf1891 Y GGCC R R CCGG Y	3	889	4166	4243	
Bfm I C TRYA G G AYRT C	2	1622	4339		
Bgl I GCCN NNN NGGC CGGN NNN NCCG	2	3151	5149		
Bli49 I GGTCTC CCAGAG	3	4477	4515	5434	
Blp I GCTNAGC CGANTCG	1	1790			
Bmc142 I RGC GCY YCG CGR	2	1804	3906		
BmcTI TGATCA ACTAGT	1	11			
Bpl I GAGNNNNNCTC CTCNNNNNGAG	4	1269	1280	3187	3198
Bpm I CTGGAGNNNNNNNNNNNNNN GACCTCNNNNNNNNNNNNNN	2	1189	3006		
Bpu10 I CC TNA GC GG ANT CG	2	1628	3259		
Bpu1268 I CCTNNNNNAGG GGANNNNNTCC	2	338	1443		
Bso I GGTCTCN NNNN CCAGAGN NNNN	3	4484	4522	5429	

Table 2 cont'd

Wednesday, July 22, 1998 2:20 PM
rmA RE sites Map (1 > 5519) 484 Cut Sites Site Summary by Enzyme

Enzyme	Freq	Position(s)				
Bso XI ACNNNNNCTCC TGNNNNNGAGG	2	2562	3809			
BsoO I CG RY CG GC YR GC	5	875	892	4012	4169	4588
BsoA I YAC GTR RTG CAY	4	680	1226	2647	2785	
BsoG I GWGCWC CWCGWG	4	1001	1558	3123	4540	
BsoK I CTTAAC GAATTG	1	2866				
BsoM I GAATG CN CTTAC GN	1	3518				
Bsb I CAACAC GTTGTG	2	1067	2558			
BscJ I CCANNNNNNTGG GGTNNNNNNACC	3	1123	1406	4799		
Bse59 I GGTNACC CCANTGG	1	1499				
BseMI GCAATG CGTTAC	3	368	4585	4978		
BseR I GAGGAGNNNNNNNN NN CTCCTCNNNNNNNN NN	3	2565	4510	4553		
Bsg I GTGCAGNNNNNNNNNNNNNN CACGTCNNNNNNNNNNNNNN	1	4208				
BshL I GATATC CTATAG	2	2368	2649			
BsiHKA I G WGCW C C WCGW G	4	1006	1563	3128	4545	
BsmBI CGTCTCN NNNN GCAGAGN NNNN	1	6355				
BsmG I TGTACA ACATGT	1	1388				
BsmH I RGCGCY YCGCGR	2	1801	3903			
BsoD I CGGCCG GCCGGC	2	888	4165			
BsoJI GCCGGC CGGCCG	1	3968				
Bsp117 I GRGCYC CYCGRG	4	227	922	1001	2670	

Table 2 Cont'd

Wednesday, July 22, 1998 2:20 PM
mA RE sites Map (1 > 5519) 484 Cut Sites Site Summary by Enzyme

Enzyme	Freq	Position(s)			
Bsp120 I G GGCC C C CCGG G	2	923	2671		
Bsp19I C CATG G G GTAC C	2	1407	2953		
Bsp24 I GACNNNNNNTGGNNNNNNN CTGNNNNNNACCNNNNNNN	3	3247	4491	4508	
Bsp24 I' NNNNN NNNNNNNNGACNN NNNNN NNNNNNNNCTGNN	3	3278	4459	4476	
Bsp8 II CTGAAG GACTTC	4	1538	3802	3995	4787
Bsp87 I CACGTG GTGCAC	3	677	1223	2644	
BspC I CTGGAC GACCTG	2	329	3492		
BspH I T CATG A A GTAC T	1	1474			
BspKT5I CTGAAGNNNNNNNNNNNNNN GACTTCNNNNNNNNNNNNNN	4	1560	3824	4017	4809
BspM I ACCTGCNNNN NNNN TGGACGNNNN NNNN	4	1524	2950	4230	4619
BsrD I GCAATG NN CGTTAC NN	3	376	4593	4970	
BsrE I CTCTTC GAGAAG	3	3	998	4992	
BsrFI R CCGG Y Y GGCC R	3	500	3969	4007	
BsrG I T GTAC A A CATG T	1	1387			
BsrW I GGATC CCTAG	3	1524	1835	4851	
BssS I C TCGT G G AGCA C	1	1065			
Bst1107 I GTA TAC CAT ATG	1	1242			
Bst29 I CCTNAGG GGANTCC	2	3609	4184		

Table 2 Cont'd

Wednesday, July 22, 1998 2:20 PM
rmA RE sites Map (1 > 5519) 484 Cut Sites Site Summary by Enzyme

Enzyme	Freq	Position(s)				
BstE II ↓ G GTNAC C C CANTG G ↑	1	1500				
BstHPI ↓ GTT AAC CAA TTG ↑	1	2869				
BstX I ↓ CCAN NNNN NTGG GGTN NNNN NACC ↑	3	1131	1414	4807		
BstZ2 I ↓ GACNNNNNNGTC CTGNNNNNNCAG ↑	4	1185	1196	5051	5062	
Bsu36 I ↓ CC TNA GG GG ANT CC ↑	2	3611	4186			
Cfr10 I ↓ R CCGG Y Y GGCC R ↑	3	500	3969	4007		
Cfr13 I YGGCCR RCCGGY	3	888	4165	4242		
Cfr9I ↓ C CCGG G G GGCC C ↑	1	1378				
CfrJ4I ↓ CCC GGG GGG CCC ↑	1	1380				
Chu II GTYRAC CARYTG	1	2886				
Dra I ↓ TTT AAA AAA TTT ↑	5	1924	1965	2028	3344	4824
Drd I ↓ GACNN NN NNGTC CTGNN NN NNCAG ↑	2	4620	4866			
Drd II GAACCA CTTGGT	4	1668	2713	3068	4858	
Dsa VI GTMKAC CAKMTG	3	1239	1584	5075		
Eae I ↓ Y GGCC R R CCGG Y ↑	3	889	4166	4243		
Ear I ↓ CTCTTCN NNN GAGAAGN NNN ↑	2	994	4988			
EcoI ↓ G GTNAC C C CANTG G ↑	1	1500				
Eci I TCCGCC AGGCGG	1	101				

Table 2 Cont'd

Wednesday, July 22, 1998 2:20 PM
rmA RE sites Map (1 > 5519) 484 Cut Sites Site Summary by Enzyme

Enzyme	Freq	Position(s)				
EclA I TACGTA ATGCAT	1	2782				
EclE I GGGCCC CCCGGG	2	922	2670			
Ecl137 I GAGCTC CTCGAG	1	1001				
EclHK I GACNN N NNGTC CTGNN N NNCAG	2	1191	5057			
Eco24I G RGCY C C YCGR G	4	232	927	1006	2675	
Eco31I GGTCTCN NNNN CCAGAGN NNNN	3	4484	4522	5429		
Eco50 I GGYRCC CCRYGG	3	846	4754	5507		
Eco52 I C GGCC G G CCGG C	2	889	4166			
Eco57 I CTGAAGNNNNNNNNNNNNNNN GACTTCNNNNNNNNNNNNNNN	4	1560	3824	4017	4809	
Eco84I G GYRC C C CRYG G	3	847	4755	5508		
Eco72 I CAC GTG GTG CAC	3	680	1226	2647		
Eco82 I GAATTC CTTAAG	2	671	2418			
Eco88I C YCGR G G RGCY C	5	1378	2469	3918	4260	5144
EcoD I TTANNNNNNNGTCY AATNNNNNNNCAGR	4	118	1609	1896	3262	
EcoD XXI TCANNNNNNNRTTC AGTNNNNNNNYAAG	3	1370	4242	5459		
EcoOR2 TCANNNNNNGTCG AGTNNNNNNCAGC	1	3011				
EcoE I GAGNNNNNNNATGC CTCNNNNNNNTACG	1	205				
EcoICR I GAG CTC CTC GAG	1	1004				

Table 2 cont'd

Wednesday, July 22, 1998 2:20 PM
rmA RE sites Map (1 > 5519) 484 Cut Sites Site Summary by Enzyme

Enzyme	Freq	Position(s)
EcoN I ↓ CCTNNN N NNAGG GGANN N NNTCC ↑	2	343 1448
Eco0109 I ↓ RG GNC CY YC CNG GR ↑	4	923 2671 3263 4796
EcoP15 I CAGCAGNNNNNNNNNNNNN GTCGTCNNNNNNNNNNNNN	2	380 547
EcoR I ↓ G AATT C C TTAA G ↑	2	672 2419
EcoR V ↓ GAT ATC CTA TAG ↑	2	2371 2852
EcoR124 I ↓ GAANNNNNNRTCG CTTNNNNNNYAGC ↑	2	1326 4272
EcoR124 II ↓ GAANNNNNNNRTCG CTTNNNNNNNYAGC ↑	4	1350 3623 4135 4195
EcoRD2 ↓ GAANNNNNNRTTC CTTNNNNNNYAAG ↑	2	3542 5288
EcoV111 ↓ A AGCT T T TCGA A ↑	3	78 2333 4600
Ecoprr I ↓ CCANNNNNNNRTGC GGTNNNNNNNYACG ↑	1	644
Esp18 I CGTCTC GCAGAG	1	5360
Esp3 I ↓ CGTCTCN NNNN GCAGAGN NNNN ↑	1	5355
Fbll ↓ GT MK AC CA KM TG ↑	3	1241 1586 5077
Fsp I ↓ TGC GCA ACG CGT ↑	1	368
Fsu I GACNNNGTC CTGNNNCAG	3	323 1128 4484
Gdi II ↓ CGGCC R GCCGG Y ↑	4	893 894 4170 4171
Gsp I CAGCTG GTCGAC	2	1791 4021

Table 2 cont'd

Wednesday, July 22, 1998 2:20 PM
rnA RE sites Map (1 > 5519) 484 Cut Sites Site Summary by Enzyme

Enzyme	Freq	Position(s)				
Hae I WGG CCW WCC GGW	3	415	3047	4245		
Hae II R GCGC Y Y CGCG R	2	1806	3908			
HaII G AATT C C TTAA G	2	672	2419			
Hga I GACGCNNNNN NNNNN CTGCGNNNNN NNNNN	5	395	759	4411	4944	4978
HgiCI G GYRC C C CRYG G	3	847	4755	5508		
HgiE II ACCNNNNNNGGT TGGNNNNNNCCA	2	3110	3871			
HinB I GRCGYC CYGCRG	2	1185	4401			
HinJCI GTY RAC CAR YTG	1	2869				
Hinc II GTY RAC CAR YTG	1	2869				
Hind III A AGCT T T TCGA A	3	78	2333	4600		
Hpo I GTT AAC CAA TTG	1	2869				
Hsp92 I GR CG YC CY GC RG	2	1187	4403			
Lsp1270 I RCATGY YGTACR	3	50	213	942		
M. SmaDam GATATC CTATAG	2	2388	2649			
Mlu1106 I RGGWCCY YCCWGGR	2	3261	4794			
Mlu113 I CC GC GG GG CG CC	2	524	4306			
MscI TGG CCA ACC GGT	1	4245				
Msl I CAYNN NNRTG GTRNN NNYAC	4	1412	4069	4239	4761	

Table 2 cont'd

Wednesday, July 22, 1998 2:20 PM
mA RE sites Msp (1 > 5519) 484 Cut Sites Site Summary by Enzyme

Enzyme	Freq	Position(s)				
Msp20 I ↓ TGGCCA ACCGGT ↑	2	4242	4248			
MspA1 I ↓ CMG CKG GKC GMC ↑	5	525	1794	1829	4024	4307
Noe I ↓ GCC GGC CGG CCG ↑	1	3971				
Nco I ↓ C CATG G G GTAC C ↑	2	1407	2953			
NgoM I ↓ G CCGG C C GGCC G ↑	1	3969				
Nhe I ↓ G CTAG C C GATC G ↑	1	3988				
NlaIII ↓ C YCGR G G RGCY C ↑	5	1382	2473	3922	4264	5148
Nru I ↓ TCG CGA AGC GCT ↑	1	1349				
Nsp I ↓ R CATG Y Y GTAC R ↑	3	55	218	947		
Pfl1108 I TCGTAG AGCATC	2	3173	5080			
PinAI ↓ A CCGG T T GGCC A ↑	1	4007				
PpeI ↓ G GGCC C C CCGG G ↑	2	927	2675			
Ppu1253 I GACGTC CTGCAG	1	1185				
Ppu8 I YACGTR RTGCAY	4	677	1223	2644	2782	
PpuM I ↓ RG GWC CY YC CWG GR ↑	2	3263	4796			
PshA I ↓ GACNN NNGTC CTGNN NNCAG ↑	1	4406				
Psp1406 I ↓ AA CG TT TT GC AA ↑	2	3582	4445			

Table 2 cont'd

Wednesday, July 22, 1998 2:20 PM
rmA RE sites Map (1 > 5519) 484 Cut Sites Site Summary by Enzyme

Enzyme	Freq	Position(s)			
PspA I C↓CCGG G G GGCC↑C	1	1378			
Pss I RG GNC↓CY YC↑CNG GR	4	926	2674	3268	4799
Pvu II CAG↓CTG GTC↑GAC	2	1794	4024		
Rhc I TCATGA AGTACT	1	1473			
RleA I CCCACANNNNNNNNN↓NNN GGGTGTNNNNNNNNN↑NNN	2	3670	5472		
Sac I G↓AGCT C C TCGA↑G	1	1006			
Sac II CC↓GC GG GG CG↑CC	2	526	4308		
Sap I GCTCTTCN↓NNN CGAGAAGN↑NNN	1	994			
SauLPI GCC↓GGC CGG↑CCG	1	3971			
Sca I AGT↓ACT TCA↑TGA	2	655	2721		
Sfc I C↓TRYA G G AYRT↑C	2	1622	4339		
SgrA I CR↓CCGG YG GY↑GGCC RC	1	3969			
Sma I CCC↓GGG GGG↑CCC	1	1380			
SmiI C↓TYRA G G ARYT↑C	3	3100	3908	5168	
Sno I GTATAC CATATG	1	1239			
SnoB I TAC↓GTA ATG↑CAT	1	2785			
Spe I A↓CTAG T T GATC↑A	1	3364			

Table 2 cont'd

Wednesday, July 22, 1998 2:20 PM
rmA RE sites Map (1 > 5519) 484 Cut Sites Site Summary by Enzyme

Enzyme	Freq	Position(s)				
Sph I G CATG C C GTAC G	1	218				
Sso I G AATT C C TTAA G	2	872	2419			
Ssp I AAT ATT TTA TAA	3	363	1928	3854		
Sst I G AGCT C C TCGA G	1	1006				
Stu I AGG CCT TCC GGA	1	915				
StySJ GACNNNNNNGTRC CTGNNNNNNCAYG	1	1412				
StySK I CGATNNNNNNNGTTA GCTANNNNNNNCAAT	1	873				
StySP I AACNNNNNNGTRC TTGNNNNNNCAYG	1	2455				
Syn II GAANNNNTTC CTTNNNNAAG	3	410	1368	3430		
Taq II GACCGANNNNNNNNN NN CTGGCTNNNNNNNNN NN	3	2720	3007	4856		
Taq II CACCCANNNNNNNNN NN GTGGGTNNNNNNNNN NN	1	5088				
Tth111 I GACN N NGTC CTGN N NCAG	3	327	1132	4488		
Tth111 II CAARCANNNNNNNNN NN. GTTYGTNNNNNNNNN NN.	3	109	2647	4740		
Ubo1220 I CCCGGG GGGCCC	1	1377				
Ubo1221 I GCTNAGC CGANTCG	2	1788	1795			
Ubo1303 I CGRYCG GCYRGC	5	871	888	4008	4165	4562
Ubo1326 I RGGNCCY YCCNGGR	4	921	2669	3261	4794	
Ubo1382 I GAATGC CTTACG	1	3511				

Table 2 cont'd

Wednesday, July 22, 1998 2:20 PM
mA RE sites Map (1 > 5519) 484 Cut Sites Site Summary by Enzyme

Enzyme	Freq	Position(s)		
Von91 I ↓ CCAN NNN NTGG GGTN NNN NACC ↑	1	2612		
Vsp I ↓ AT TA AT TA AT TA ↑	1	1890		
Xcm I ↓ CCANNNN N NNNNTGG GGTNNNN N NNNNACC ↑	1	1164		
Xma I ↓ C CCGG G G GGCC C ↑	1	1378		
XmaIII ↓ C GGCC G G CCGG C ↑	2	889	4168	
Xmn I ↓ GAANN NNTTC CTTNN NNAAG ↑	3	415	1373	3435

Table 2 cont'd

List of restriction enzymes (alphabetical order), cutting *H. influenzae* Rd *rm*B 5 times or less, with positions of restriction sites indicated.

Aat II G ACGT C C TGCA G	1	1189				
Aca I TTCGAA AAGCTT	1	1721				
Aca III TGCGCA ACGCGT	1	364				
Acc I GT MK AC CA KM TG	3	1240	1585	4831		
Acc III T CCGG A A GGCC T	2	1296	3511			
Ace II G CTAG C C GATC G	1	3746				
Ace III CAGCTCNNNNNNNN NNNN GTCGAGNNNNNNNN NNNN	3	1073	3542	4579		
Acr I CYCGRG GRGCYC	5	1378	2222	3671	4013	4897
AcsI R AATT Y Y TTAA R	5	671	2173	3185	3951	4115
Afe24R I GCCGGC CGGCCG	1	3722				
Afl III A CRYG T T GYRC A	3	678	1222	2398		
Afl IV AGTACT TCATGA	2	651	2472			
Age I A CCGG T T GGCC A	1	3761				
Alw I GGATCNNNN N CCTAGNNNN N	2	1532	4400			
AlwN I CAG NNN CTG GTC NNN GAC	2	1045	4246			
Amo I TCGCGA AGCGCT	1	1345				
Aos III CCGCGG GGCGCC	2	521	4058			
Apo I G GGCC C C CCGG G	2	926	2429			

Table 2 Cont'd

Wednesday, July 22, 1998 2:39 PM
rmB Map (1 > 5273) 505 Cut Sites

Site Summary by Enzyme

Enzyme	Freq	Position(s)				
Apo I R AATT Y Y TTAA R	5	671	2173	3185	3951	4115
AquI C YCGR G G RGCY C	5	1377	2223	3872	4014	4898
Asp52 I AASCTT TTCGAA	3	76	2086	4353		
Asp5H I GCATGC CGTACG	1	212				
Asp78 I AGGCCT TCCGGA	2	411	1683			
Ate I CCATGG GGTACC	2	1405	2706			
AtuC I TGATCA ACTAGT	1	10				
Avo I C YCGR G G RGCY C	5	1377	2223	3672	4014	4898
Avr II C CTAG G G GATC C	2	620	1687			
Bae I ACNNNNGTAYC TGNNNNCATRG	2	787	4384			
Bae I NNNNNNNNNNNNNNNACNNN NNNNNNNNNNNNNNNTGNNN	2	813	4410			
Bal I TGG CCA ACC GGT	1	3999				
Ban I G GYRC C C CRYG G	3	846	4509	5262		
Ban II G RGCY C C YCGR G	4	231	926	1005	2429	
BavI CAG CTG GTC GAC	1	3778				
Bbi7911 I TCCGGA AGGCCT	2	1295	3510			
Bbr I A AGCT T T TCGA A	3	77	2087	4354		
Bbs I GAAGACNN NNNN CTTCTGNN NNNN	3	1547	3321	5032		
Bca I GCGC CGCG	5	365	1100	3134	3858	3720

Table 2 cont'd

Wednesday, July 22, 1998 2:38 PM
rmB Map (1 > 5273) 505 Cut Sites

Site Summary by Enzyme

Enzyme	Freq	Position(s)			
Bce83 I CTTGAGNNNNNNNNNNNNN GAACTCNNNNNNNNNNNNN	3	2875	3883	4907	
Bcg I GCANNNNNNTCGNNNNNNN CGTNNNNNNAGCNNNNNNN	1	1071			
Bcg I' NN NNNNNNNNNNGCANNN NN NNNNNNNNNNCGTNNN	1	1037			
Bcl I T GATC A A CTAG T	1	11			
Bco102 II GAAGAC CTTCTG	3	1539	3313	5024	
Bco163 I CTRYAG GAYRTC	1	4092			
Bco35 I CTGGAG GACCTC	2	1166	2738		
Bcu I ACTAGT TGATCA	2	3117	3123		
BfiBBI Y GGCC R R CCGG Y	3	888	3920	3997	
Bfm I C TRYA G G AYRT C	1	4093			
Bgl I GCCN NNN NGGC CGGN NNN NCCG	2	2905	4903		
Bli49 I GGTCTC CCAGAG	3	4231	4269	5188	
Bme142 I RGC GCY YCG CGR	1	3660			
BmeT I TGATCA ACTAGT	1	10			
Bpl I GAGNNNNNCTC CTCNNNNNGAG	4	1268	1279	2941	2952
Bpm I CTGGAGNNNNNNNNNNNNN GACCTCNNNNNNNNNNNNN	2	1188	2760		
Bpu10 I CC TNA GC GG ANT CG	1	3013			
Bpu1268 I CCTNNNNNAGG GGANNNNNTCC	2	337	1442		

Table 2 cont'd

Wednesday, July 22, 1998 2:38 PM
rmB Map (1 > 5273) 505 Cut Sites

Site Summary by Enzyme

Enzyme	Freq	Position(s)				
Bso I ↓ CGTCTCN NNNN GCAGAGN NNNN ↑	3	4238	4276	5183		
Bso XI ↓ ACNNNNNCTCC TGNNNNNGAGG ↑	2	2316	3563			
BscO I ↓ CG RY CG GC YR GC ↑	5	874	891	3786	3923	4320
BscA I ↓ YAC GTR RTG CAY ↑	4	679	1225	2401	2539	
BscG I GWGCWC CWCGWG	4	1000	1557	2877	4294	
BscK I GTTAAC CAATTG	1	2620				
BscM I ↓ GAATG CN CTTAC GN ↑	1	3272				
Bsb I CAACAC GTTGTG	2	1086	2312			
BscJ I CCANNNNNNTGG GGTNNNNNNACC	3	1122	1405	4553		
Bse59 I GGTNACC CCANTGG	2	1498	1710			
BseMI ↓ CCAATG GGTTAC ↑	3	367	4339	4730		
BseR I ↓ GAGGAGNNNNNNNN NN CTCCTCNNNNNNNN NN ↑	3	2319	4264	4307		
Bsq I GTGCAGNNNNNNNNNNNNNN CACGTCNNNNNNNNNNNNNN	1	3962				
BshL I GATATC CTATAG	2	2122	2403			
BsiHKA I ↓ G WGCW C C WCGW G ↑	4	1005	1562	2882	4299	
BsmBI ↓ CGTCTCN NNNN GCAGAGN NNNN ↑	1	5109				
BsmG I TGTACA ACATGT	1	1385				
BseH I RGCGCY YCGCGR	1	3657				
BsoD I CGGCCG GCCGGC	2	887	3919			

Table 2 cont'd

Wednesday, July 22, 1998 2:38 PM
rmB Map (1 > 5273) 505 Cut Sites

Site Summary by Enzyme

Enzyme	Freq	Position(s)			
BsoJI GCCGGC CGGCCG	1	3722			
Bsp117 I GRGCYC CYCGRG	4	228	921	1000	2424
Bsp120 I G GGCC C C CCGG G	2	922	2425		
Bsp19I C CATG G G GTAC C	2	1406	2707		
Bsp24 I GACNNNNNNTGGNNNNNNN CTGNNNNNNNACCNNNNNNNN	3	3001	4245	4262	
Bsp24 I NNNNN NNNNNNNNGACNN NNNNN NNNNNNNNCTGNN	3	3033	4213	4230	
Bsp6 II CTGAAG GACTTC	4	1537	3556	3749	4541
Bsp87 I CACGTG GTGCAC	3	676	1222	2398	
BspG I CTGGAC GACCTG	2	328	3246		
BspH I T CATG A A GTAC T	1	1473			
BspKT5I CTGAAGNNNNNNNNNNNNNN GACTTCNNNNNNNNNNNNNN	4	1559	3578	3771	4563
BspLU11 II TCTAGA AGATCT	2	1678	1684		
BspM I ACCTGCNNNN NNNN TGGACGNNNN NNNN	4	1523	2704	3984	4373
BsrD I GCAATG NN CGTTAC NN	3	375	4347	4724	
BsrE I CTCTTC GAGAAG	3	2	897	4748	
BsrFI R CCGG Y Y GGCC R	3	499	3723	3761	
BsrG I T GTAC A A CATG T	1	1388			
BsrW I GGATC CCTAG	2	1523	4405		

Table 2 cont'd

Enzyme	Freq	Position(s)				
BssS I C TCGT G G AGCA C	1	1084				
Bst1107 I GTA TAC CAT ATG	1	1291				
Bst29 I CCTNAGG GGANTCC	2	3383	3838			
BstE II G GTNAC C C CANTG G	2	1499	1711			
BstHPI GTT AAC CAA TTG	1	2623				
BstX I CCAN NNNN NTGG GGTN NNNN NACC	3	1130	1413	4581		
BstZ2 I GACNNNNNGTC CTGNNNNNCAG	4	1184	1195	4805	4816	
Bsu36 I CC TNA GG GG ANT CC	2	3365	3940			
Cfo I G CG C C GC G	5	368	1103	3137	3861	3723
Cfr10 I R CCGG Y Y GGCC R	3	499	3723	3761		
Cfr13 I YGGCCR RCCGGY	3	887	3919	3996		
Cfr9I C CCGG G G GGCC C	1	1377				
CfrJ9I CCC GGG GGG CCC	1	1379				
Chu II GTYRAC CARYTG	1	2620				
Csp45 I TT CG AA AA GC TT	1	1723				
Dra I TTT AAA AAA TTT	3	1782	3098	4578		
Drd I GACNN NN NNGTC CTGNN NN NNCAG	2	4374	4620			

Table 2 cont'd

Wednesday, July 22, 1998 2:38 PM
rmB Map (1 > 5273) 505 Cut Sites

Site Summary by Enzyme

Enzyme	Freq	Position(s)				
Drd II GAACCA CTTGGT	3	2467	2822	4412		
Dsa V! GTMKAC CAKMTG	3	1238	1583	4829		
Eae I Y GGCC R R CCGG Y	3	888	3920	3997		
Ear I CTCTTCN NNN GAGAAGN NNN	2	993	4742			
EcoI G GTNAC C C CANTG G	2	1499	1711			
Ecl I TCCGCC AGGCGG	1	100				
EclA I TACGTA ATGCAT	1	2536				
EclE I GGGCCC CCCGGG	2	921	2424			
Ecl137 I GAGCTC CTCGAG	1	1000				
EclHK I GACNN N NNGTC CTGNN N NNCAG	2	1190	4811			
Eco24I G RGCY C C YCGR G	4	231	926	1005	2429	
Eco31I GGTCTCN NNNN CCAGAGN NNNN	3	4238	4276	5183		
Eco50 I GGYRCC CCRYGG	3	845	4508	5261		
Eco52 I C GGCC G G CCGG C	2	888	3920			
Eco57 I CTGAAGNNNNNNNNNNNNNNN GACTTCNNNNNNNNNNNNNNN	4	1559	3578	3771	4563	
Eco64I G GYRC C C CRYG G	3	846	4509	5262		
Eco72 I CAC GTG GTG CAC	3	679	1225	2401		
Eco82 I GAATTC CTTAAG	2	670	2172			
Eco88I C YCGR G G RGCY C	5	1377	2223	3672	4014	4898

Table 2 contd

Wednesday, July 22, 1998 2:38 PM
mB Map (1 > 5273) 505 Cut Sites

Site Summary by Enzyme

Enzyme	Freq	Position(s)			
EcoD I ↓ TTANNNNNNNGTCY AATNNNNNNNCAGR ↑	3	117	1757	3018	
EcoD XXI ↓ TCANNNNNNNRTTC AGTNNNNNNNYAAG ↑	3	1389	3996	5213	
EcoDR2 ↓ TCANNNNNNGTCG AGTNNNNNNCAGC ↑	1	2765			
EcoE I ↓ GAGNNNNNNNATGC CTCNNNNNNNTACG ↑	1	204			
EcoICR I ↓ GAG CTC CTC GAG ↑	1	1003			
EcoN I ↓ CCTNN N NNAGG GGANN N NNTCC ↑	2	342	1447		
EcoO109 I ↓ RG GNC CY YC CNG GR ↑	4	922	2425	3017	4550
EcoP15 I CAGCAGNNNNNNNNNNNNN GTCGTCNNNNNNNNNNNNN	2	379	548		
EcoR I ↓ G AATT C C TTAA G ↑	2	671	2173		
EcoR V ↓ GAT ATC CTA TAG ↑	2	2125	2406		
EcoR124 I ↓ GAANNNNNNRTCG CTTNNNNNNYAGC ↑	2	1325	4026		
EcoR124 II ↓ GAANNNNNNNRTCG CTTNNNNNNNYAGC ↑	4	1349	3377	3889	3949
EcoRD2 ↓ GAANNNNNNNRTTC CTTNNNNNNNYAAG ↑	2	3296	5040		
EcoVIII ↓ A AGCT T T TCGA A ↑	3	77	2087	4354	
Ecoprr I ↓ CCANNNNNNNRTGC GGTNNNNNNNYACG ↑	1	643			
Esp1B I CGTCTC GCAGAG	1	5114			

Table 2 contd

Wednesday, July 22, 1998 2:38 PM
rmB Map (1 > 5273) 505 Cut Sites

Site Summary by Enzyme

Enzyme	Freq	Position(s)				
Esp3 I CGTCTCN NNNN GCAGAGN NNNN	1	5109				
Fbl I GT MK AC CA KM TG	3	1240	1585	4831		
Fsp I TGC GCA ACG CGT	1	387				
Fsu I GACNNNGTC CTGNNNCAG	3	322	1127	4238		
Gdi II CGCCC R GCCGG Y	4	892	893	3924	3925	
Gsp I CAGCTG GTCGAC	1	3775				
Hae I WGG CCW WCC GGW	4	414	1686	2801	3999	
Hae II R GCGC Y Y CGCG R	1	3882				
HaII G AATT C C TTAA G	2	671	2173			
HgiCI G GYRC C C CRYG G	3	848	4509	5262		
HgiE II ACCNNNNNNNGGT TGGNNNNNNNCCA	2	2864	3625			
Hha I G CG C C GC G	5	368	1103	3137	3681	3723
HinB I GRCGYC CYGCRG	3	1184	1737	4155		
HinJCI GTY RAC CAR YTG	1	2823				
HinPI I G CG C C GC G	5	368	1101	3135	3659	3721
Hinc II GTY RAC CAR YTG	1	2823				
Hind III A AGCT T T TCGA A	3	77	2087	4354		

Table 2 cont'd

Wednesday, July 22, 1998 2:38 PM
rmB Map (1 > 5273) 505 Cut Sites

Site Summary by Enzyme

Enzyme	Freq	Position(s)				
Kpn I GTT AAC CAA TTG	1	2623				
Hsp92 I GR CG YC CY GC RG	3	1186	1739	4157		
Lsp1270 I RCATGY YGTACR	3	49	212	941		
M. SssDam GATATC CTATAG	2	2122	2403			
Mlu1108 I RGGWCCY YCCWGGR	2	3015	4548			
Mlu113 I CC GC GG GG CG CC	2	523	4060			
Nsc I TGG CCA ACC GGT	1	3999				
Nsi I CAYNN NNRTG GTRNN NNYAC	4	1411	3823	3993	4515	
Nsp2O I TGGCCA ACCGGT	2	3996	4002			
NspA1 I CMG CKG GKC GMC	3	524	3778	4081		
Noe I GCC GGC CGG CCG	1	3725				
Nco I C CATG G G GTAC C	2	1406	2707			
NgoM I G CCGG C C GGCC G	1	3723				
Nhe I G CTAG C C GATC G	1	3742				
NI1387/7 I C YCGR G G RGCY C	5	1381	2227	3876	4018	4902
Nru I TCG CGA AGC GCT	1	1348				
Nsp I R CATG Y Y GTAC R	3	54	217	946		

Table 2 cont'd

Wednesday, July 22, 1998 2:38 PM
mnB Map (1 > 5273) 505 Cut Sites

Site Summary by Enzyme

Enzyme	Freq	Position(s)			
Pf11108 I TCGTAG AGCATC	2	2927	4834		
PinAI A CCGG T T GGCC A	1	3761			
PpeI G GCCC C C CCGG G	2	926	2429		
Ppu1253 I GACGTC CTGCAG	1	1184			
Ppu6 I YACGTR RTGCAY	5	676	1222	2398	2536
PpuM I RG GWC CY YC CWG GR	2	3017	4550		
PshA I GACNN NNGTC CTGNN NNCAG	1	4180			
Psp1406 I AA CG TT TT GC AA	4	1614	1628	3336	4199
PspAI C CCGG G G GGCC C	1	1377			
Pss I RG GNC CY YC CNG GR	4	925	2428	3020	4553
Pvu II CAG CTG GTC GAC	1	3778			
Rhc I TCATGA AGTACT	1	1472			
RleA I CCCACANNNNNNNNN NNN GGGTGTNNNNNNNNN NNN	2	3424	5226		
Soc I G AGCT C C TCGA G	1	1005			
Soc II CC GC GG GG CG CC	2	525	4062		
Sap I GCTCTTCN NNN CGAGAAGN NNN	1	993			
SouLPI GCC GGC CGG CCG	1	3725			

Table 2 cont'd

Wednesday, July 22, 1998 2:38 PM
rmB Map (1 > 5273) 505 Cut Sites

Site Summary by Enzyme

Enzyme	Freq	Position(s)			
Sco I ↓ AGT ACT TCA TGA ↑	2	654	2475		
Sfc I ↓ C TRYA G G AYRT C ↑	1	4093			
SgrA I ↓ CR CCGG YG GY GGCC RC ↑	1	3723			
Sim I ↓ GGGTC CCCAG ↑	4	2597	2689	4080	5184
Sma I ↓ CCC GGG GGG CCC ↑	1	1379			
SmlI ↓ C TYRA G G ARYT C ↑	3	2854	3862	4922	
Sno I GTATAC CATATG	1	1238			
SnoB I ↓ TAC GTA ATG CAT ↑	1	2539			
Spe I ↓ A CTAG T T GATC A ↑	1	3118			
Sph I ↓ G CATG C C GTAC G ↑	1	217			
SsoI ↓ G AATT C C TTAA G ↑	2	671	2173		
Ssp I ↓ AAT ATT TTA TAA ↑	2	362	3408		
SstI ↓ G AGCT C C TCGA G ↑	1	1005			
Stu I ↓ AGG CCT TCC GGA ↑	2	414	1686		
StySJ ↓ GAGNNNNNNGTRC CTCNNNNNNCAYG ↑	1	1411			
StySKI ↓ CGATNNNNNNNGTTA GCTANNNNNNNCAAT ↑	1	872			

Table 2 cont'd

Wednesday, July 22, 1998 2:38 PM
rmB Map (1 > 5273) 505 Cut Sites

Site Summary by Enzyme

Enzyme	Freq	Position(s)				
StySP I AACNNNNNNGTRC TTGNNNNNNCAYG	1	2209				
Syn II GAANNNNTTC CTTNNNNAAG	3	409	1367	3184		
Taq II GACCGANNNNNNNNN NN CTGGCTNNNNNNNNN NN	3	2474	2761	4610		
Taq II CACCCANNNNNNNNN NN GTGGGTNNNNNNNNN NN	1	4842				
Tth111 I GACN N NGTC CTGN N NCAG	3	326	1131	4242		
Tth111 II CAARCANNNNNNNNN NN. GTTYGTNNNNNNNNN NN.	3	108	2401	4494		
Ubo1220 I CCCGGG GGGCCC	1	1376				
Ubo1303 I CGRYCG GCYRGC	5	870	887	3762	3919	4318
Ubo1326 I RGGNCCY YCCNGGR	4	920	2423	3015	4548	
Ubo1382 I GAATGC CTTACG	1	3265				
VanB1 I CCAN NNN NTGG GGTN NNN NACC	1	2366				
Xba I T CTAG A A GATC T	1	1679				
Xcm I CCANNNN N NNNNTGG GGTNNNN N NNNNACC	1	1163				
Xma I C CCGG G G GGCC C	1	1377				
XmaIII C GGCC G G CCGG C	2	888	3920			
Xmn I GAANN NNTTC CTTNN NNAAG	3	414	1372	3189		

CLAIMS

We claim:

1. A method for bacterial speciation, comprising:

i) isolating bacterial DNA from a sample, said DNA comprising DNA encoding 16S and 23S rRNA;

ii) digesting said isolated DNA with one or more restriction enzymes under conditions such that restriction fragments are produced, said restriction fragments comprising a first digestion product of said DNA encoding 16S and 23S rRNA, said first digestion product comprising at least a portion of said DNA encoding 16S rRNA and at least a portion of said DNA encoding 23S rRNA;

iii) separating said restriction fragments;

iv) detecting of said first digestion product; and

v) comparing said digestion product with signature bands of one or more bacterial species.

2. The method of Claim 1, wherein said detecting comprising reacting a probe with said digestion product under conditions such that said probe hybridizes to siad first digestion product.

3. A method for bacterial speciation, comprising:

i) isolating bacterial DNA from a sample, said DNA comprising DNA encoding 16S and 23S rRNA;

ii) digesting said isolated DNA with one or more restriction enzymes under conditions such that restriction fragments are produced, said restriction fragments comprising first and second digestion products of said DNA encoding 16S and 23S rRNA, said first digestion product being larger than said second digestion product, and comprising at least a portion of said DNA encoding 16S rRNA and at least a portion of said DNA encoding 23S rRNA;

iii) separating of said restriction fragments;

iv) detecting said first and second digestion products; and

v) comparing said digestion products with signature bands of one or more bacterial species.

4. A method for bacterial speciation, comprising:

a) providing i) a first biological sample comprising bacterial DNA from a known bacterial species, and ii) a second biological sample comprising bacterial DNA from a bacterium whose species is unknown;

b) isolating i) a first preparation of bacterial DNA from said first sample and ii) a second preparation of bacterial DNA from said second sample, said DNA of said first and second preparations comprising DNA encoding 16S and 23S rRNA;

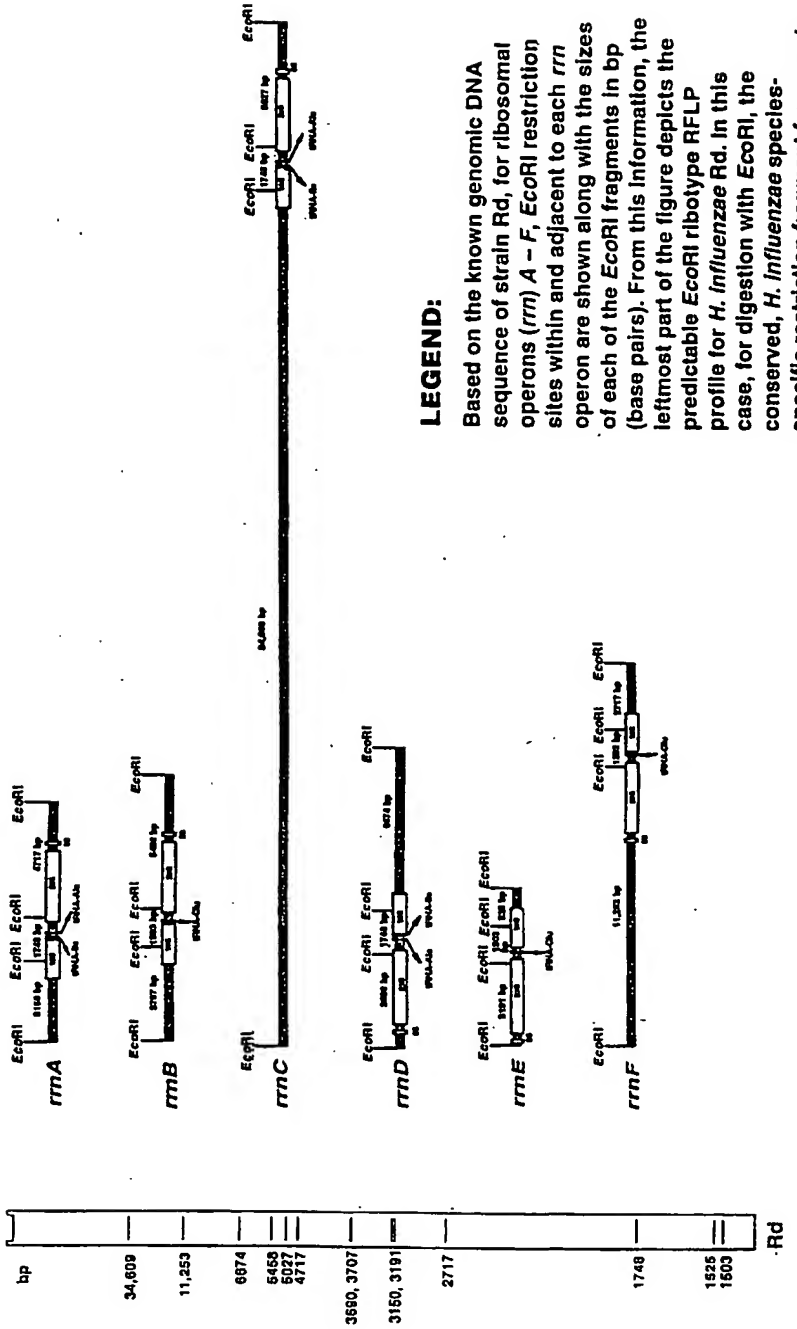
c) digesting, in any order, i) said first preparation of isolated DNA with one or more restriction enzymes under conditions such that a first preparation of restriction fragments are produced, said first preparation of restriction fragments comprising a first digestion product, said first digestion product comprising at least a portion of said DNA encoding 16S rRNA and at least a portion of said DNA encoding 23S rRNA, and ii) said second preparation of isolated DNA with one or more restriction enzymes under conditions such that a second preparation of restriction fragments are produced, said second preparation of restriction fragments comprising a second digestion product, said second digestion product comprising at least a portion of said DNA encoding 16S rRNA and at least a portion of said DNA encoding 23S rRNA;

d) separating, in any order, i) said restriction fragments from said first preparation, and ii) said restriction fragments from said second preparation; and

e) comparing of said first and second digestion products.

FIG. 1

The 6 ribosomal RNA operons of the genomically sequenced *H. influenzae* strain Rd

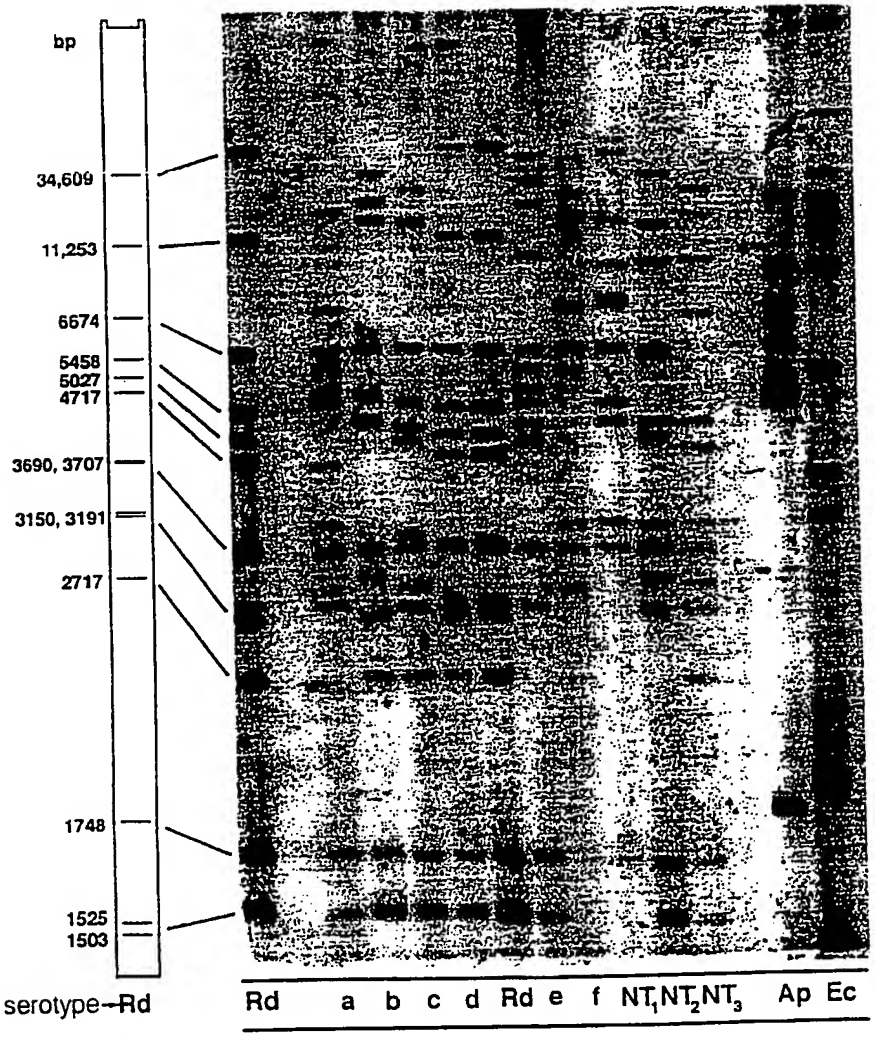


LEGEND:

Based on the known genomic DNA sequence of strain Rd, for ribosomal operons (*rrn*) *A – F*, *Eco*RI restriction sites within and adjacent to each *rrn* operon are shown along with the sizes of each of the *Eco*RI fragments in bp (base pairs). From this information, the leftmost part of the figure depicts the predictable *Eco*RI ribotype RFLP profile for *H. influenzae* Rd. In this case, for digestion with *Eco*RI, the conserved, *H. influenzae* species-specific restriction fragment from each *rrn* operon is that defined by the two internal *Eco*RI cutting sites.

FIG. 2

EcoRI ribotype RFLPs of *H. influenzae* isolates from diverse sources, including the genomically sequenced strain Rd



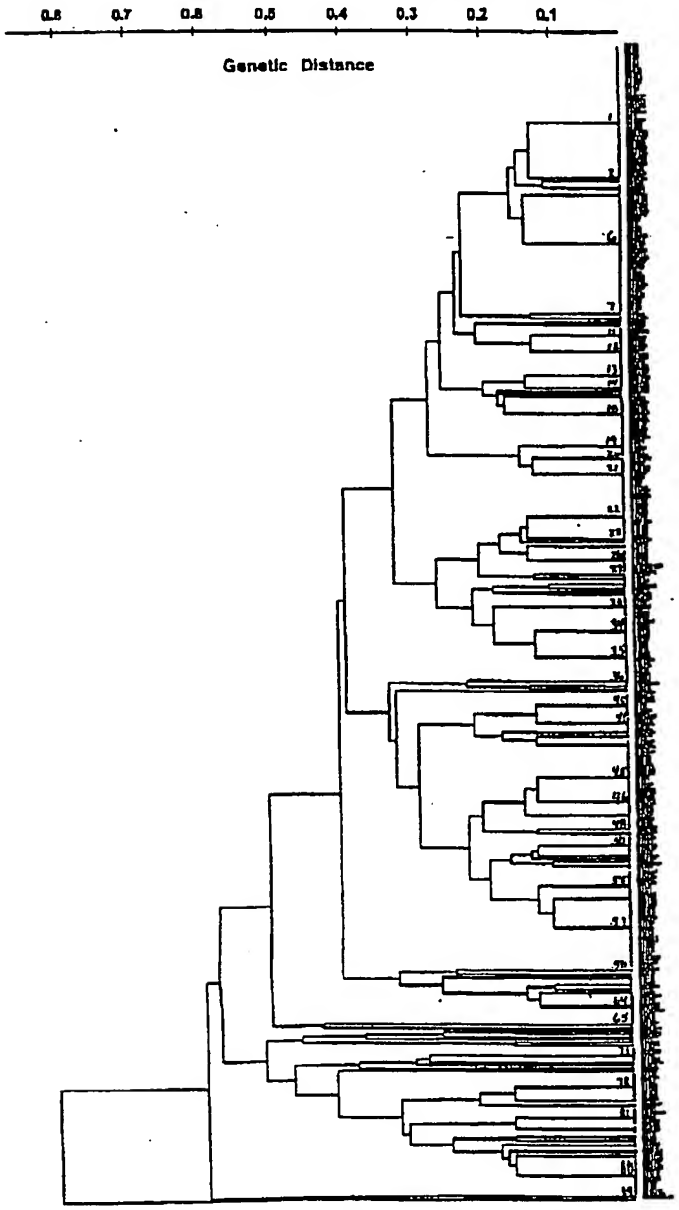
LEGEND.

Leftmost portion of the figure depicts the predictable *Eco*RI ribotype RFLP profile of the genomically *sequenced H. influenzae* strain Rd. Actual RFLP profiles of this strain appears in lanes 1 and 6 (so labeled). Other *H. influenzae* isolates, as indicated, are serotypes a, c, c, d, e, f and NT (non-typable, i.e. unencapsulated). *Eco*RI ribotype RFLPs of 2 non-*H. influenzae* isolates are shown in the farthestmost right lanes, that for *A. pleuroneumonia* (Ap) and *E. coli* (Ec). Conserved, species specific *H. influenzae Eco*RI ribotype bands can be seen at 1,748 and 1,503 bp (base pairs). These 2 fragments are absent from the type f isolate and one NT (NT_1), both of which have subsequently been found by 16S *rrn* gene sequencing to actually not be members of the species *H. influenzae*.

FIG. 3

EcoRI Ribotype-based Phylogenic Tree of A Diverse Collection of *H. influenzae* isolates (types a – f, and non-typable) from Variant Clinical and Environmental Sources and Geographical Locales

(unpublished results)



LEGEND.

Phylogeny displayed here for >400 independent isolates has been independently confirmed based on serotyping, capsule operon gene polymorphism analysis, and comparative DNA sequencing of 16S *rrn* and neutral genes.

FIG. 4

DNA sequence and restriction map of the *H. influenzae* Rd *rrn*A operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less.

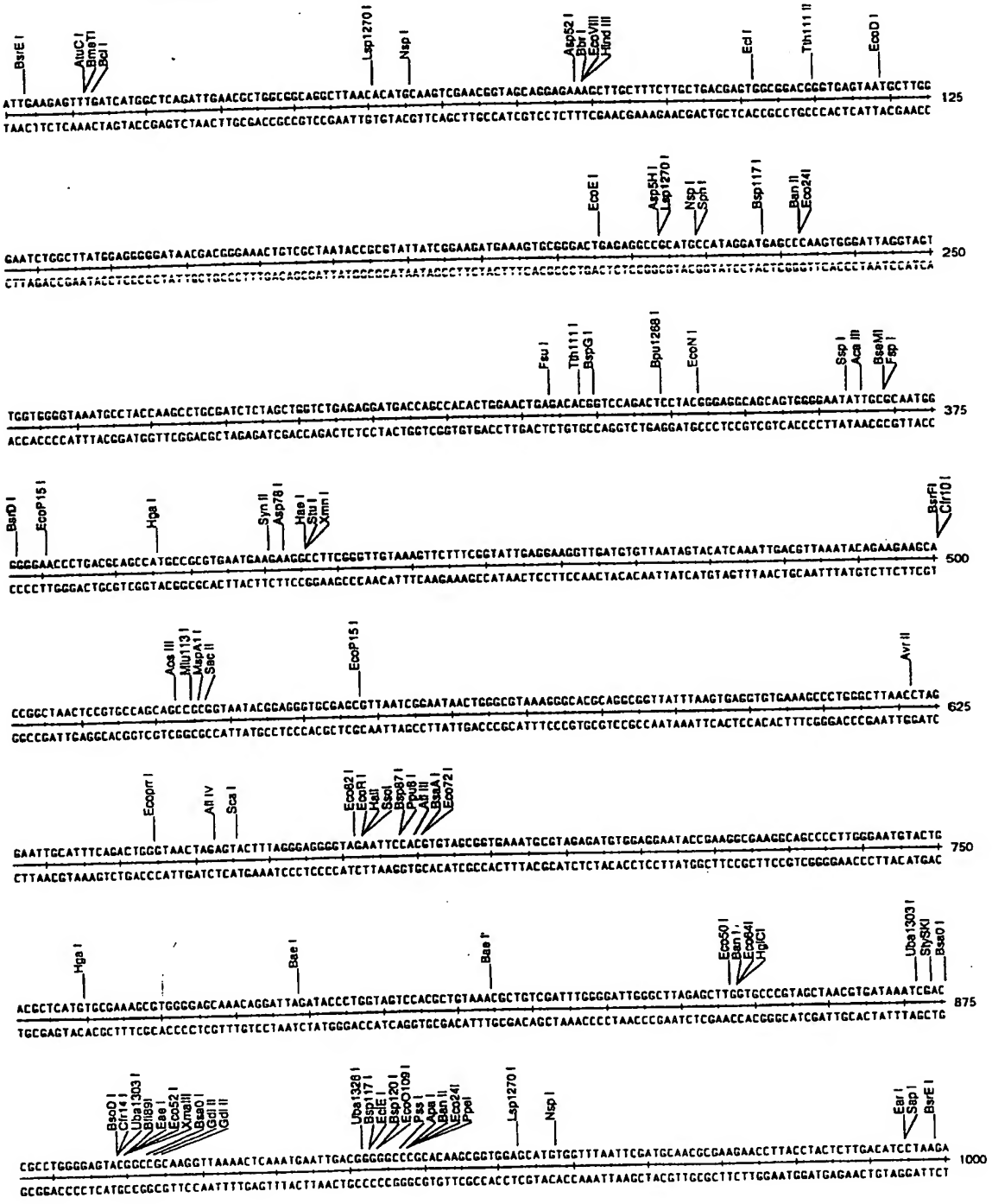


FIG. 4 cont'd

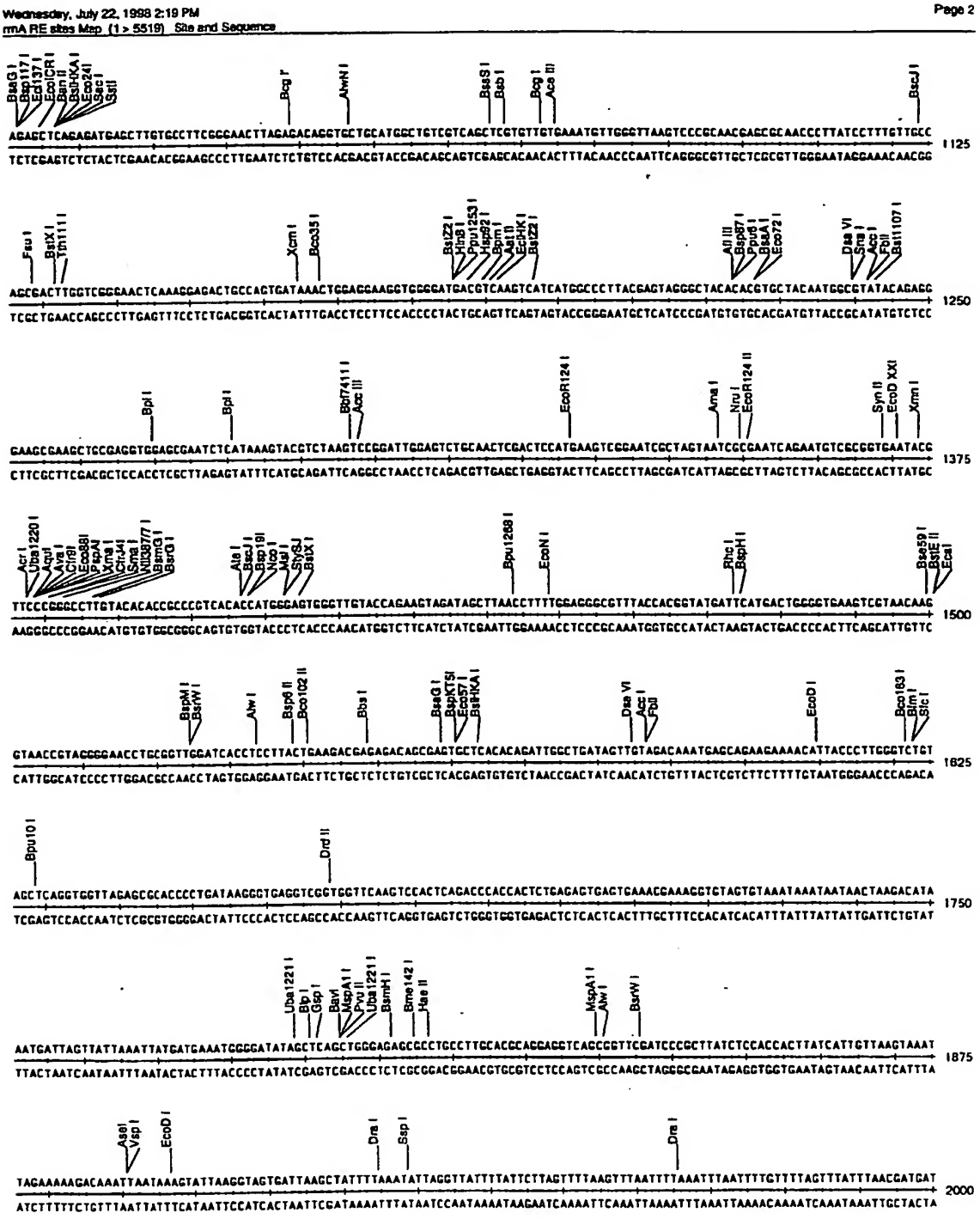

Wednesday, July 22, 1998 2:19 PM
Page 2
mA RE sites Map (1 > 5519) Site and Sequence
BsaG I
Bsp117 I
Ed137 I
EcoICR I
Ban II
BspHKA I
Eco24 I
Sac I
SstI
Bcg I
AlwN I
BssS I
Bsb I
Bcg I
Ace III
BscJ I
AGAGCTCAGAGATGAGCTTGTGCCTTCGGGAACTTAGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGCC 1125
TCTCGAGTCTCTACTCGAACACGGAAGCCCTTGAATCTCTGTCCACGACGTACCGACAGCAGTCGAGCACAACACTTTACAACCCAATTCAGGGCGTTGCTCGCGTTGGGAATACGAAACAACGG
Fau I
BstX I
Tth111 I
Xcm I
Bco35 I
BsiZ2 I
HinS1 I
Ppu1253 I
Hsp92 I
Bpm I
Aat II
EcoHK I
BstZ2 I
Afl III
Bsp87 I
Ppu6 I
BsaA I
Eco72 I
Dsa VI
Sna I
Acc I
FbiI
Bst1107 I
AGCGACTTGGTCGGCAACTCAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTCAAGTCATCATGGCCCTTACGAGTAGGGCTACACACGTGCTACAATGGCGTATACAGAGG 1250
TCGCTGAACCAGCCCTTGAGTTTCCTCTGACGGTCACTATTTGACCTCCTTCCACCCCTACTGCAGTTCAGTAGTACCGGGAATGCTCATCCCGATGTGTGCACGATGTTACCGCATATGTCTCC
BpI I
BpI I
Bbl7411 I
Acc III
EcoR124 I
Ama I
Nru I
EcoR124 II
Sym II
EcoD XXI
Xmn I
GAAGCGAAGCTGCGAGGTGGAGCGAATCTCATAAAGTACGTCTAAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGCGAATCAGAATGTCGCGGTGAATACG 1375
CTTCGCTTCGACGCTCCACCTCGCTTAGAGTATTTCATGCAGATTCAGGCCTAACCTCAGACGTTGAGCTGAGGTACTTCAGCCTTAGCGATCATTAGCGCTTAGTCTTACAGCGCCACTTATGC
Acr I
Uba1220 I
AquI
Ava I
CfrBI
Eco88I
PspAI
Xma I
ChrJ4I
Sma I
NIU3677 I
BsmG I
BsrG I
Ale I
BscLI
Bsp19I
Nco I
MaI I
ShyBJ
BstX I
Bpu1288 I
EcoN I
Rhc I
BspH I
Bse59 I
BstE II
EcaI
TTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGTACCAGAAGTAGATAGCTTAACCTTTTGGAGGGCGTTTACCACGGTATGATTCATGACTGGGGTGAAGTCGTAACAAG 1500
AAGGGCCCGGAACATGTGTGGCGGGCAGTGTGGTACCCTCACCCAACATGGTCTTCATCTATCGAATTGGAAAACCTCCCGCAAATGGTGCCATACTAAGTACTGACCCCACTTCAGCATTGTTC
BspM I
BsrW I
AlwI
Bsp6 II
Bco102 II
Bbs I
BsaG I
BspKTSI
Eco57 I
BstHKA I
Dsa VI
Acc I
FbiI
EcoO I
Bco183 I
BlmI
SfcI
GTAACCGTACGGGAACCTGCGGTTGGATCACCTCCTTACTGAAGACGAGAGACAGCGAGTGCTCACACAGATTGGCTGATAGTTGTAGACAAATGAGCAGAAGAAAACATTACCCTTGGGTCTGT 1625
CATTGGCATCCCCTTGGACGCCAACCTAGTGGAGGAATGACTTCTGCTCTCTGTCGCTCACGAGTGTGTCTAACCGACTATCAACATCTGTTTACTCGTCTTCTTTTGTAATGGGAACCCAGACA
Bpu10 I
DrdII
AGCTCAGGTGGTTAGAGCGCACCCCTGATAAGGGTGAGGTCGGTGGTTCAAGTCCACTCAGACCCACCACTCTCAGAGTGAGTGAAACGAAAGGTGTAGTGTAAATAAATAATAACTAAGACATA 1750
TCGAGTCCACCAATCTCGCGTGGGGACTATTCCCACTCCAGCCACCAAGTTCAGGTGAGTCTGGGTGGTGAGACTCTCACTCACTTTGCTTTCCACATCACATTTATTTATTATTGATTCTGTAT
Uba1221 I
Bp1 I
Gsp I
BavI
MspA1 I
PvuII
Uba1221 I
BsmI I
Bme142 I
Hae II
MspA1 I
AlwI
BsrW I
AATGATTAGTTATTAAATTATGATGAAATGGGGATATAGCTCAGCTGGGAGAGCGCCTGCCTTGCACGCAGGAGGTCAGCGGTTCGATCCCGCTTATCTCCACCACTTATCATTGTTAAGTAAAT 1875
TTACTAATCAATAATTTAATACTACTTTACCCCTATATCGAGTCGACCCTCTCGCGGACGGAACGTGCGTCCTCCAGTCGCCAAGCTAGGGCGAATAGAGGTGGTGAATAGTAACAATTCATTTA
AseI
VspI
EcoD I
Dra I
Bsp I
Dra I
TAGAAAAAGACAAATTAATAAAGTATTAAGGTAGTGATTAAGCTATTTTAAATATTAGGTTATTTTATTCTTAGTTTTAAGTTTAATTTTAAATTTAATTTTGTTTTAGTTTATTTAACGATGAT 2000
ATCTTTTTCTGTTTAATTATTTCATAATTCCATCACTAATTCGATAAAATTTATAATCCAATAAAATAAGAATCAAAATTCAAATTAAAATTTAAATTAAAACAAAATCAAATAAATTGCTACTA

FIG. 4 Cont'd

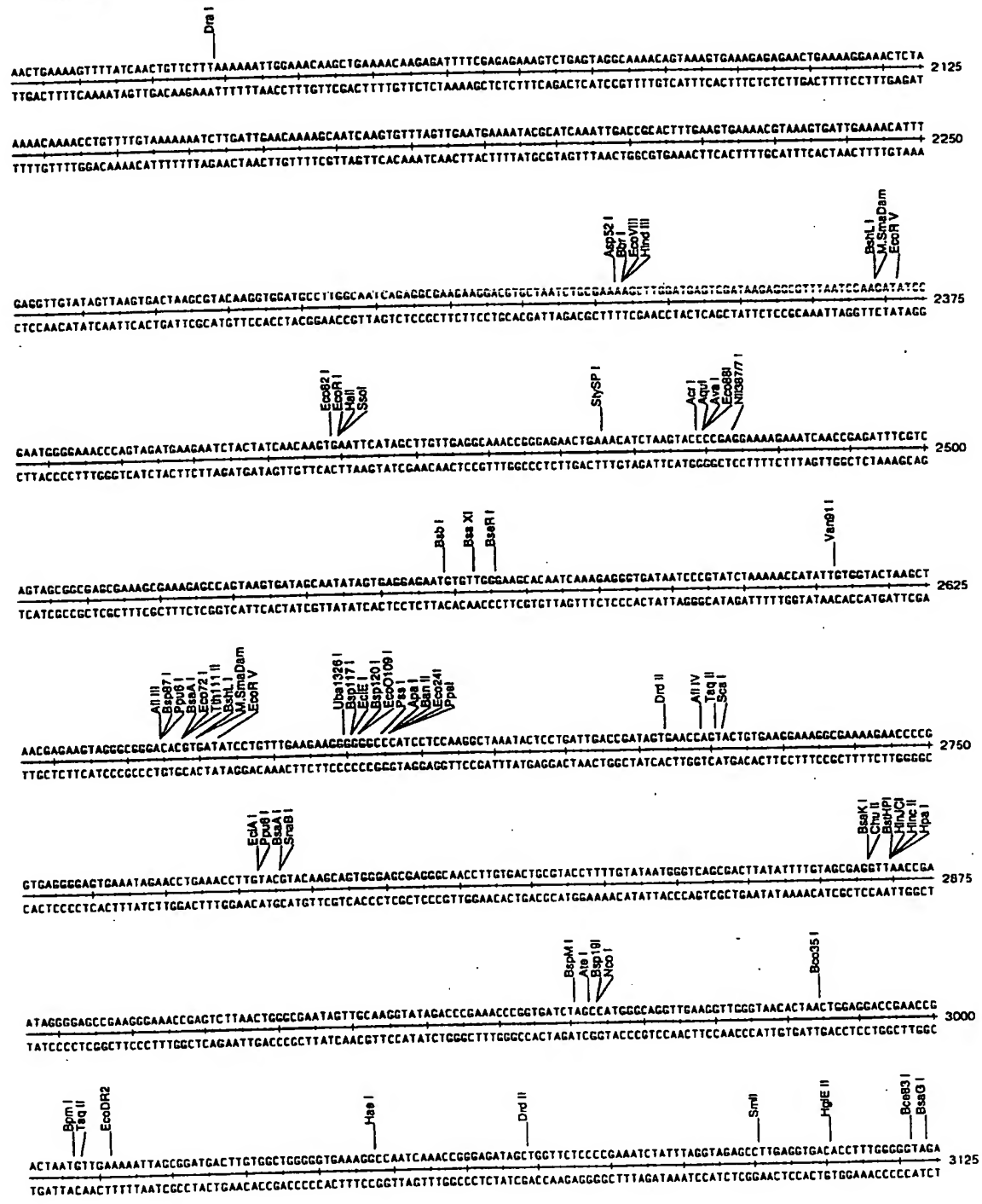

Dra I
2125
Asp52 I
Bbr I
EcoV III
Mind III
BshL I
M.SmaDam
EcoR V
2250
Eco82 I
EcoR I
Hali
SsoI
StySP I
Acc I
Aqu I
Ava I
Eco88I
NIS8777 I
2375
Bsb I
Bsa XI
BseR I
Van91 I
2500
All III
Bsp87 I
Ppu6 I
BsaA I
Eco72I
Ybh111 II
BshL I
M.SmaDam
EcoR V
Uba1326 I
Bsp1171
EclE I
Bsp120 I
EcoO109 I
Psa I
Apa I
Bsn II
Eco24I
PpeI
Drd II
All IV
Taq II
Sca I
2625
EclA I
Ppu6 I
BsaA I
SnaB I
BsaK I
Chu II
BshPI
HinJCI
Hinc II
Hpa I
2750
BspM I
Ale I
Bsp19I
Nco I
Boo35 I
2875
Bpm I
Taq II
EcoDR2
Hae I
Drd II
SmiI
HgiE II
Bce83 I
BsaG I
3000
3125

FIG. 4 Cont'd

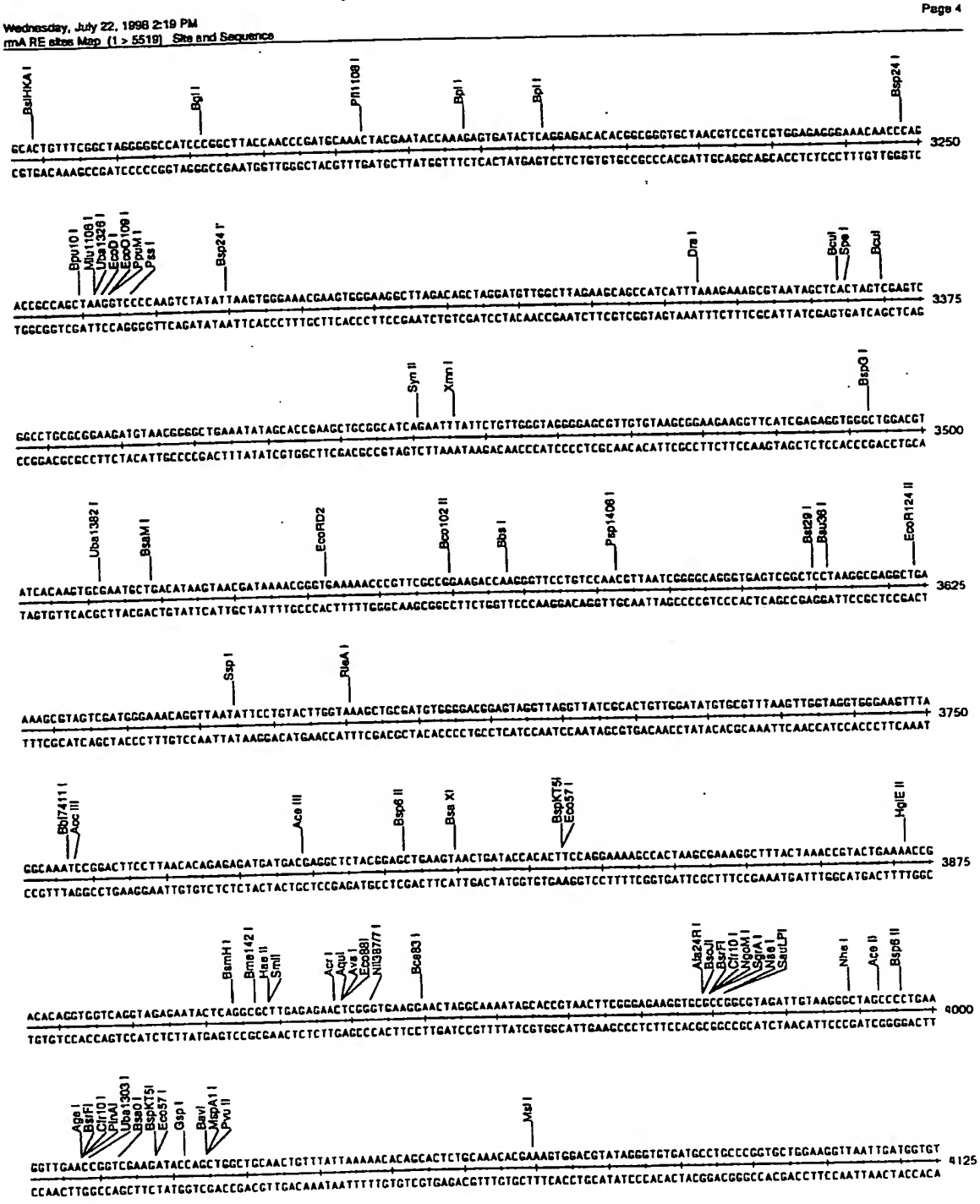

Wednesday, July 22, 1998 2:19 PM
Page 4
mA RE sites Map (1 > 5519) Site and Sequence
BstHKA I
Bgl I
Pfl1108 I
Bpl I
Bpl I
Bsp24 I
3250
Bpu10 I
Mlu1108 I
Uba1328 I
EcoO I
EcoO109 I
PpuM I
Pst I
Bsp24 I
Dra I
BcuI
Spe I
BcuI
3375
Syn II
Xmn I
BspQ I
3500
Uba1382 I
BsaM I
EcoRD2
Bco102 II
Bbs I
Psp1406 I
Bst29 I
Bau36 I
EcoR124 II
3625
Ssp I
RseA I
3750
Bbl7411 I
Acc III
Ace III
Bsp6 II
Bsa XI
BspKT5I
Eco57 I
HgiE II
3875
BamH I
Bme142 I
Hae II
SmiI
Acr I
Aqu I
Ava I
Eco88I
NI13877 I
Bcl83 I
Aba24R I
BsoJI
BsrFI
Cfr10 I
VcoM I
SgrA I
NIle I
SacLPI
Nhe I
Ace II
Bsp5 II
4000
Age I
BstFI
Cfr10 I
PinAI
Uba1303 I
Bsa0 I
BspKT5I
Eco57 I
Gsp I
BsvI
MspA1 I
Pvu II
MsI I
4125

FIG. 4 cont'd

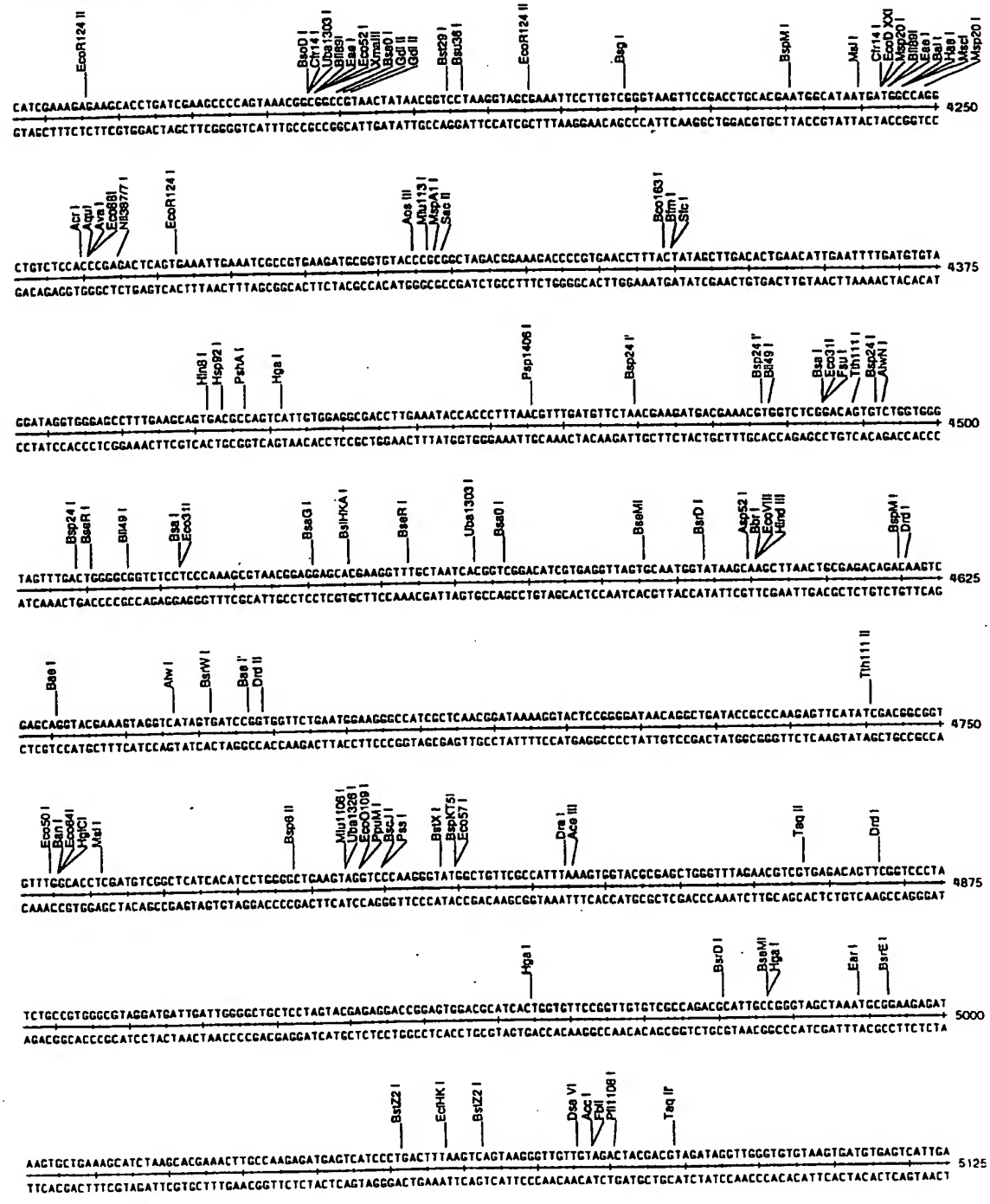

Wednesday, July 22, 1998 2:19 PM
rmA RE sites Map (1 > 5519) Site and Sequence
Page 5
EcoR124 II
BsoD I
Ch14 I
Uba1303 I
Bll89i
Eae I
Eco52 I
XmaIII
Bse0 I
Gdi II
Gdi II
Bst29 I
Bsu36 I
EcoR124 II
Bsq I
BspM I
Msl I
Ch14 I
EcoD XXI
Msp20 I
Bll89i
Eae I
Bsa I
Hae I
Msc I
Msp20 I
Acr I
Aqu I
Ava I
Eco88I
NI39377 I
EcoR124 I
Aos III
Mlu113 I
MspA1 I
Sac II
Boo163 I
Bfm I
Stc I
Hin8 I
Hsp92 I
PshA I
Hga I
Pap1406 I
Bsp24 I'
Bsp24 I'
Bll49 I
Bsa I
Eco31 I
Fsu I
Tth111 I
Bsp24 I
AlwN I
Bsp24 I
BseR I
Bll49 I
Bsa I
Eco31 I
BseG I
BslHKA I
BseR I
Uba1303 I
BsaO I
BseMI
BsrD I
Asp52 I
Bbr I
EcoVIII
Hind III
BspM I
Drd I
Bae I
Alw I
BsrW I
Bae I'
Drd II
Tth111 II
Eco50 I
Ban I
Eco64 I
HgiCI
Msl I
Bsp6 II
Mlu1108 I
Uba1326 I
EcoO109 I
PpuM I
BscJ I
Pst I
BstX I
BspKT5I
Eco57 I
Dra I
Ace III
Taq II
Drd I
Hga I
BsrD I
BseMI
Hga I
Ear I
BsrE I
BstZ2 I
EcHK I
BstZ2 I
Dse VI
Asc I
PsiI
Pfl1108 I
Taq IF

FIG. 4 Cont'd

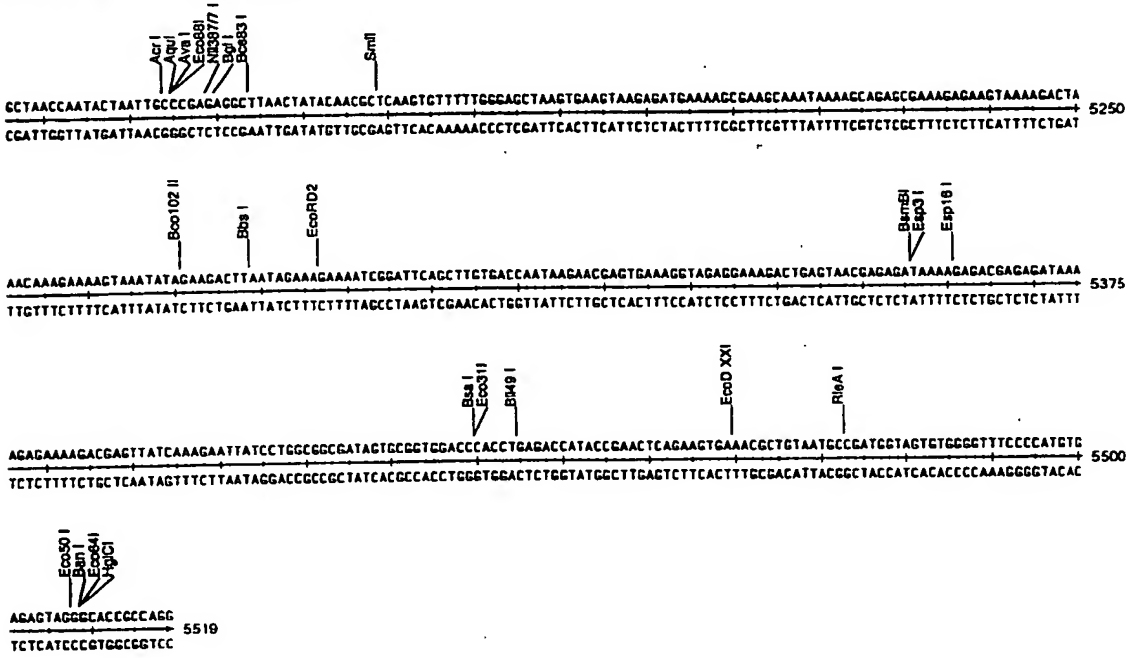

Acr I
Aqu I
Ava I
Eco88I
N339?? I
Bgl I
Bse83 I
SmI
GCTAACCAATACTAATTGCCCGAGAGCCTTAACTATACAACGCTCAAGTGTTTTTGGCAGCTAAGTGAAGTAAGAGATGAAAAGCGAAGCAAATAAAAGCAGAGCGAAAGAGAAGTAAAACACTA
CGATTGGTTATGATTAACGGGCTCTCCGAATTGATATGTTGCGAGTTCACAAAAACCCTCGATTCACTTCATTCTCTACTTTTCGCTTCGTTTATTTTCGTCTCGCTTTCTCTTCATTTTCTGAT
5250
Bco102 II
Bbs I
EcoRD2
BamBI
Esp3 I
Esp1B I
AACAAAGAAAAGTAAATATAGAAGACTTAATAGAAAGAAAATCGGATTCAGCTTGTGACCAATAAGAACGAGTGAAAGGTAGAGGAAAGACTGAGTAACGAGAGATAAAAGAGACGAGAGATAAA
TTGTTTCTTTTCATTTATATCTTCTGAATTATCTTTCTTTTAGCCTAAGTCGAACACTGGTTATTCTTGCTCACTTTCCATCTCCTTTCTGACTCATTGCTCTCTATTTTCTCTGCTCTCTATTT
5375
Bsa I
Eco31I
BtG49 I
EcoD XXI
RleA I
AGAGAAAAGACGAGTTATCAAAGAATTATCCTGGCGGCGATAGTGCGGTCGACCCACCTGAGACCATACCGAACTCAGAAGTGAAACGCTGTAATGCCCATGGTAGTGTGGGGTTTCCCCATGTG
TCTCTTTTCTGCTCAATAGTTTCTTAATAGGACCGCCGCTATCACGCCACCTGGGTGGACTCTGGTATGGCTTGAGTCTTCACTTTGCGACATTACGGCTACCATCACACCCCAAAGGGGTACAC
5500
Eco50 I
Ban I
Eco64 I
AgiCI
AGAGTAGGGCACCGCCAGG
TCTCATCCCGTGGCGGTCC
5519

FIG. 4 cont'd

DNA sequence and restriction map of the *H. influenzae* Rd *rrn*B operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less.

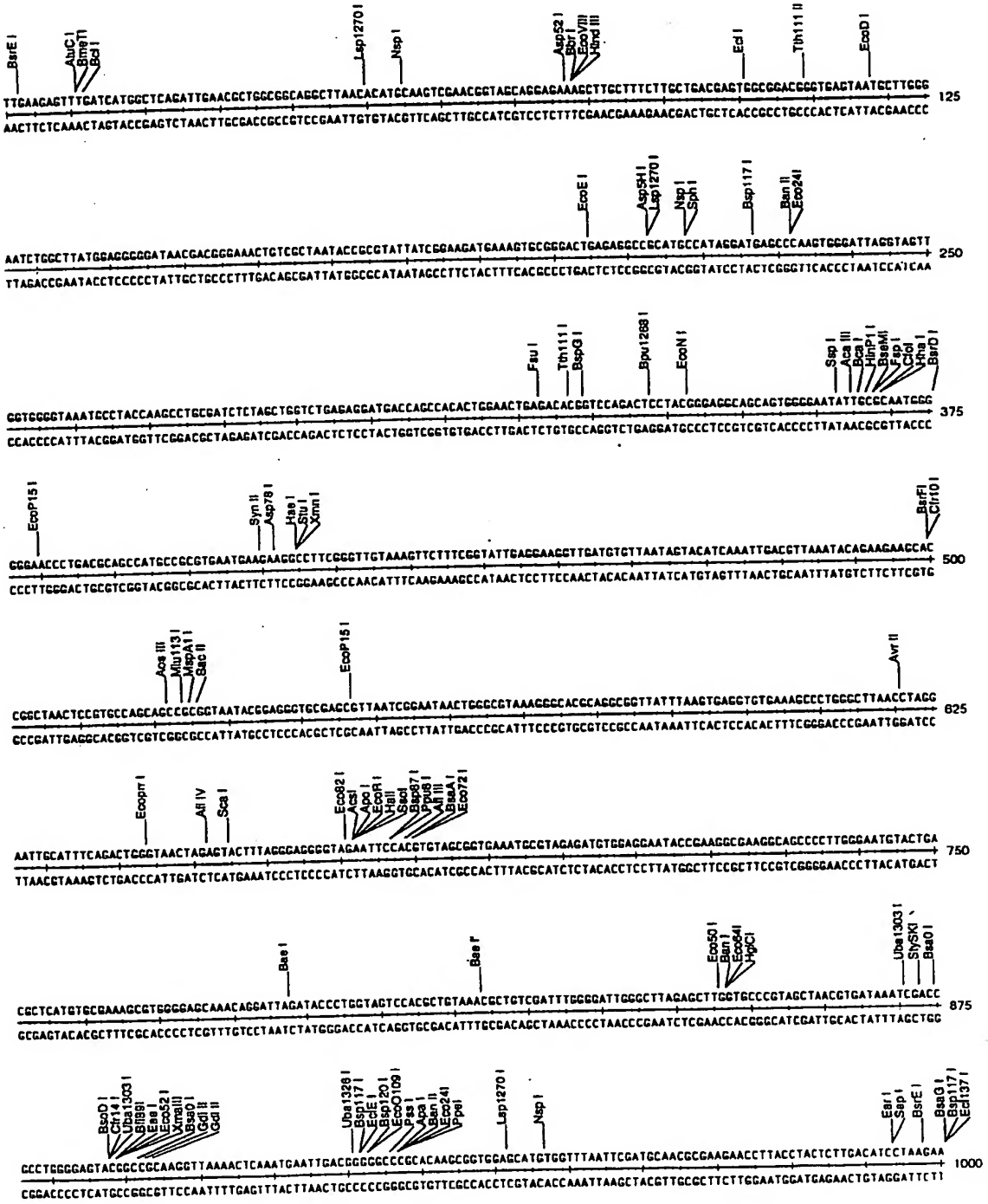


FIG. 4 cont'd

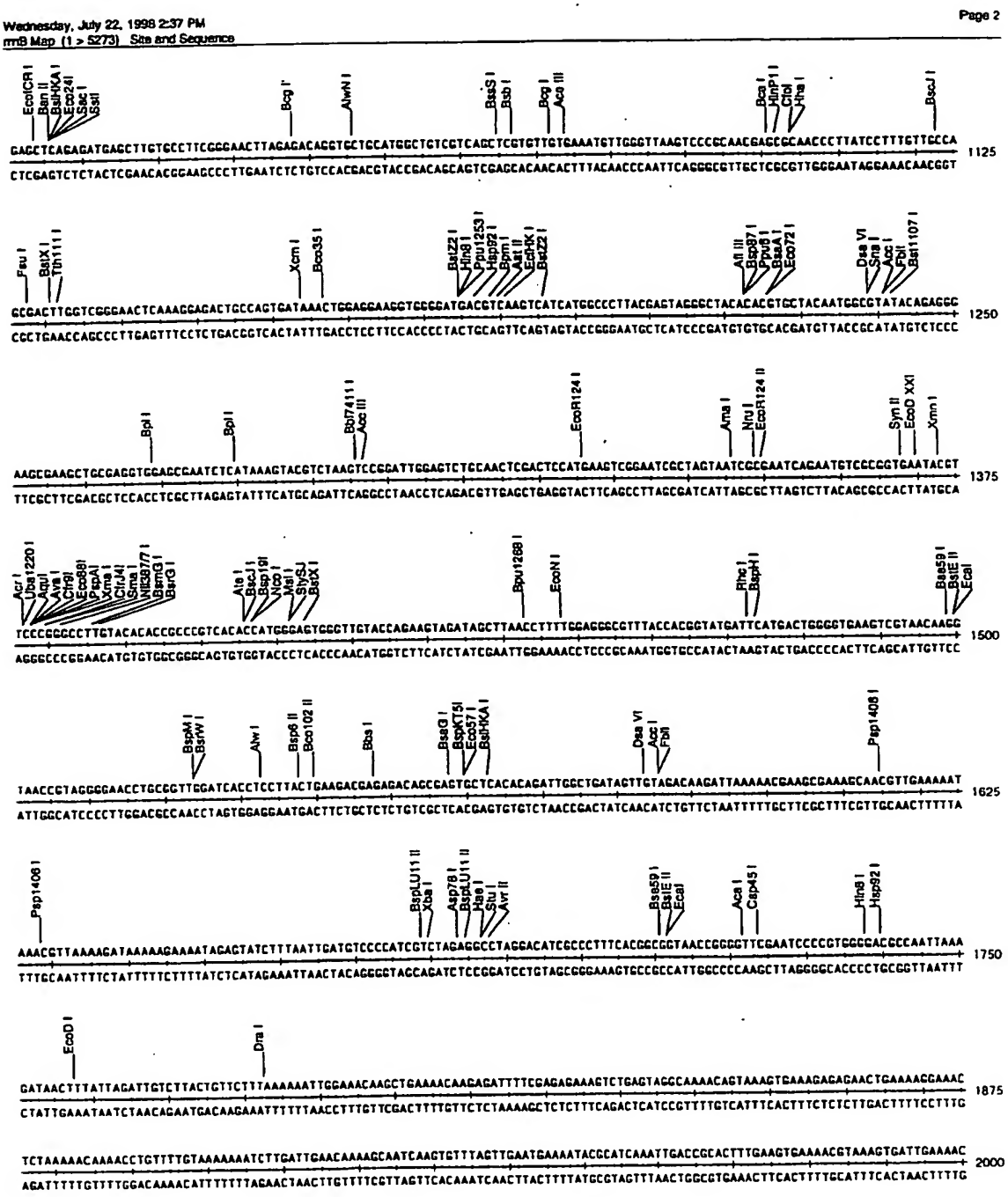

Wednesday, July 22, 1998 2:37 PM
Page 2
rrnB Map (1 > 5273) Site and Sequence

FIG. 4 cont'd

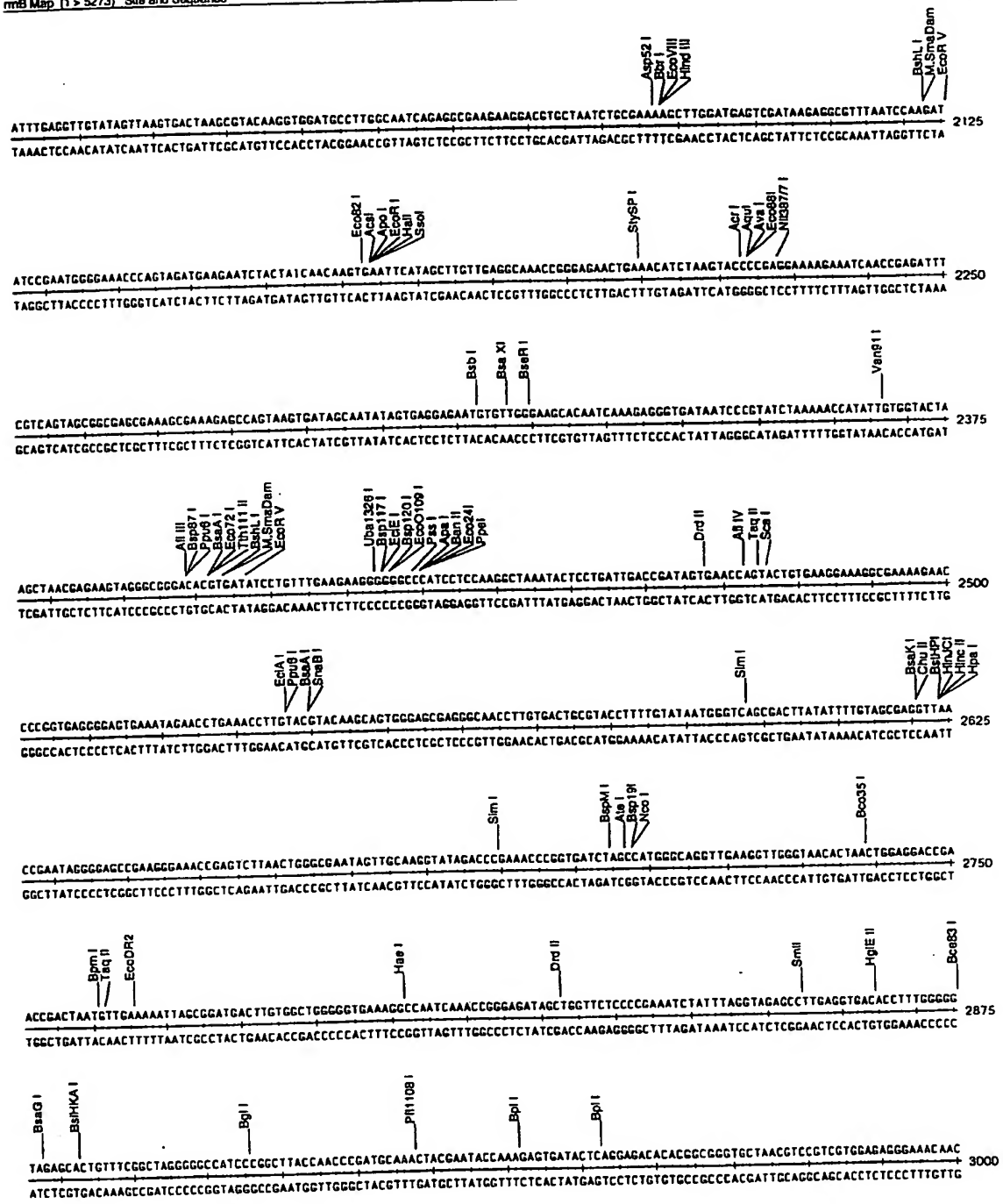

Asp52 I
Bbr I
EcoVIII
Hind III
BshL I
M.SmaDam
EcoR V
ATTTGAGGTTGTATAGTTAAGTGACTAAGCGTACAAGGTGGATGCCTTGGCAATCAGAGCGAAGAAGGACGTGCTAATCTGCGAAAAGCTTGGATGAGTCGATAAGAGGCGTTTAATCCAAGAT
TAAACTCCAACATATCAATTCACTGATTCGCATGTTCCACCTACGGAACCGTTAGTCTCCGCTTCTTCCTGCACGATTAGACGCTTTTCGAACCTACTCAGCTATTCTCCGCAAATTAGGTTCTA
2125
Eco82 I
AclI
Apo I
EcoR I
HaII
SsoI
StySP I
Acr I
Aqu I
Ava I
Eco88I
NIt3877 I
ATCCGAATGGGGAAACCCAGTAGATGAAGAATCTACTATCAACAAGTGAATTCATAGCTTGTTGAGGCAAACCGGGAGAACTGAAACATCTAAGTACCCCGAGGAAAAGAAATCAACCGAGATTT
TAGGCTTACCCCTTTGGGTCATCTACTTCTTAGATGATAGTTGTTCACTTAAGTATCGAACAACTCCGTTTGGCCCTCTTGACTTTGTAGATTCATGGGGCTCCTTTTCTTTAGTTGGCTCTAAA
2250
BsbI
Bsa XI
BseR I
Van91 I
CGTCAGTAGCGGCGAGCGAAAGCGAAAGAGCCAGTAAGTGATAGCAATATAGTGAGGAGAATGTGTTCGGAAGCACAATCAAAGAGGGTGATAATCCCGTATCTAAAAACCATATTGTGGTACTA
GCAGTCATCGCCGCTCGCTTTCGCTTTCTCGGTCATTCACTATCGTTATATCACTCCTCTTACACAACCCTTCGTGTTAGTTTCTCCCACTATTAGGGCATAGATTTTTGGTATAACACCATGAT
2375
Afl III
Bpp87 I
Fpu9 I
Bsa A I
Eco72 I
Tth111 II
BshL I
M.SmaDam
EcoR V
Uba1326 I
Bsp117 I
EclE I
Bsp120 I
EcoO109 I
Pst I
Apa I
Ban II
Eco24I
PpeI
Drd II
Alf IV
Taq II
Sca I
AGCTAACGAGAAGTAGGGCGGGACACGTGATATCCTGTTTGAAGAAGGGGGGCCCATCCTCCAAGGCTAAATACTCCTGATTGACCGATAGTGAACCAGTACTGTGAAGGAAAGGCGAAAAGAAC
TCGATTGCTCTTCATCCCGCCCTGTGCACTATAGGACAAACTTCTTCCCCCCGGGTAGGAGGTTCCGATTTATGAGGACTAACTGGCTATCACTTGGTCATGACACTTCCTTTCCGCTTTTCTTG
2500
EclA I
FpuB I
BsaA I
SnaB I
Sim I
BsaK I
Chu II
BstHPI
HinJCI
Hinc II
Hpa I
CCCGGTGAGGGGAGTGAAATAGAACCTGAAACCTTGTACGTACAAGCAGTGGGAGCGAGGGCAACCTTGTGACTGCGTACCTTTTGTATAATGGGTCAGCGACTTATATTTTGTAGCGAGGTTAA
GGGCCACTCCCCTCACTTTATCTTGGACTTTGGAACATGCATGTTCGTCACCCTCGCTCCCGTTGGAACACTGACGCATGCAAAACATATTACCCAGTCGCTGAATATAAAACATCGCTCCAATT
2625
Sim I
BspM I
Ate I
Bsp19I
Nco I
Bco35 I
CCGAATAGGGGAGCCGAAGGGAAACCGAGTCTTAACTGGGCGAATAGTTGCAAGGTATAGACCCGAAACCCGGTGATCTAGCCATGGGCAGGTTGAAGGTTGGGTAACACTAACTGGAGGACCGA
GGCTTATCCCCTCGGCTTCCCTTTGGGTCAGAATTGACCCGCTTATCAACGTTCCATATCTGGGCTTTGGGCCACTAGATCGGTACCCGTCCAACTTCCAACCCATTGTGATTGACCTCCTGGCT
2750
BpmI
Taq II
EcoDR2
Hae I
Drd II
SmlI
HgiE II
Bce83 I
ACCGACTAATGTTGAAAAATTAGCGGATGACTTGTGGCTGGGGGTGAAAGGCCAATCAAACCGGGAGATAGCTGGTTCTCCCCGAAATCTATTTAGGTAGAGCCTTGAGGTGACACCTTTGGGGG
TGGCTGATTACAACTTTTTAATCGCCTACTGAACACCGACCCCCACTTTCCGGTTAGTTTGGCCCTCTATCGACCAAGAGGGGCTTTAGATAAATCCATCTCGGAACTCCACTGTGGAAACCCCC
2875
BsaG I
BshKA I
Bgl I
PH1108 I
Bpl I
Bpl I
TAGAGCACTGTTTCGGCTAGGGGCCATCCCGGCTTACCAACCCGATGCAAACTACGAATACCAAAGAGTGATACTCAGGAGACACACGGCGGGTGCTAACGTCCGTCGTGGAGAGGGAAACAAC
ATCTCGTGACAAAGCCGATCCCCCGGTAGGGCCGAATGGTTGGGCTACGTTTGATGCTTATGGTTTCTCACTATGAGTCCTCTGTGTGCCGCCCACGATTGCAGGCAGCACCTCTCCCTTTGTTG
3000

FIG. 4 cont'd

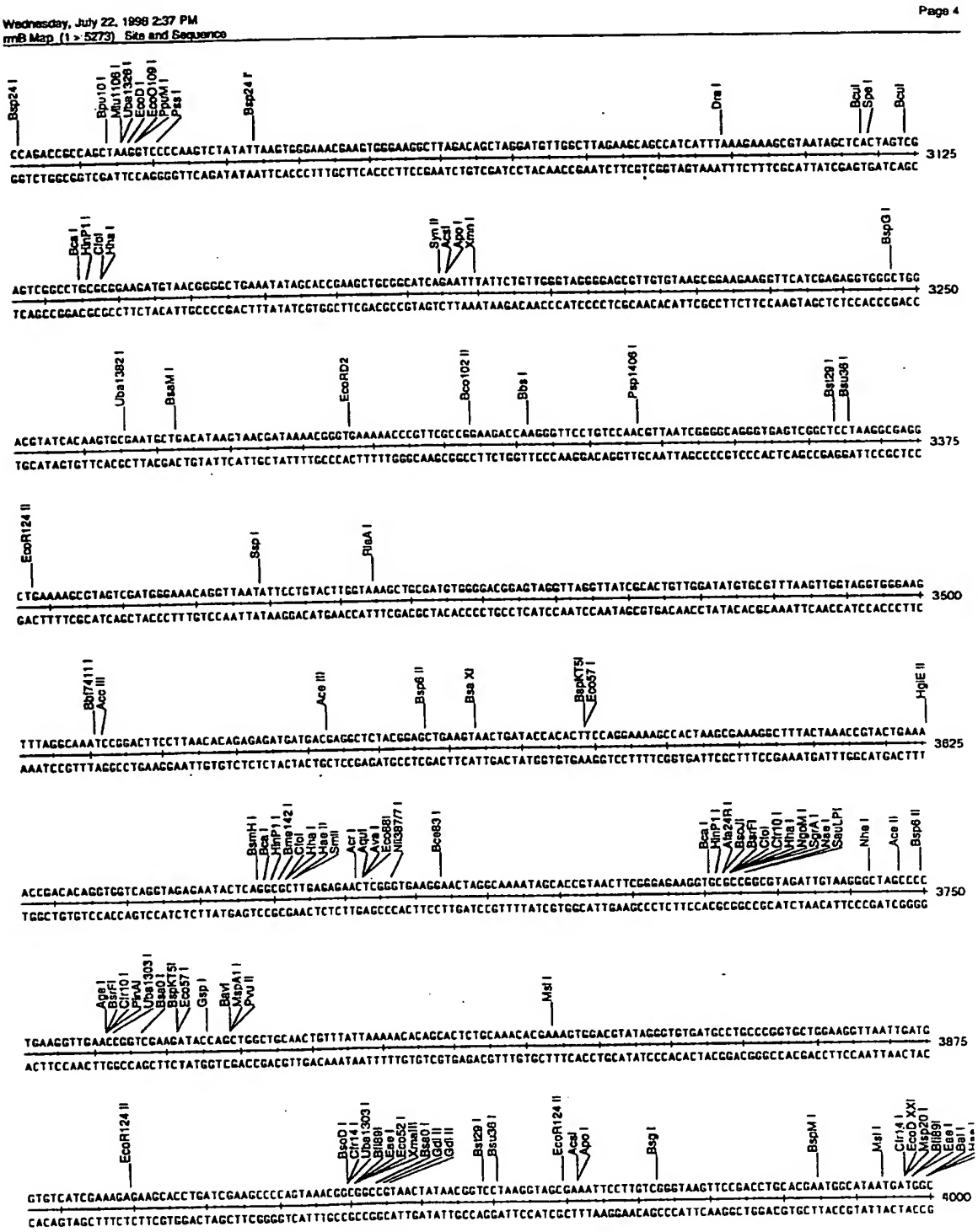

Wednesday, July 22, 1998 2:37 PM
Page 4
rmB Map (1 > 5273) Site and Sequence
Bsp24 I
Bpu10 I
Mlu1106 I
Uba1328 I
EcoD I
EcoO109 I
PpuM I
Pss I
Bsp24 II
Dra I
BcuI
Spe I
BcuI
BspG I
Sym II
AcsI
Apo I
Xmn I
Bca I
HinP1 I
ClaI
Hha I
Uba1382 I
BssM I
EcoRD2
Bco102 II
Bbs I
Psp1406 I
Bst29 I
Bsu36 I
EcoR124 II
Sap I
RigA I
Bbi7411 I
Acc III
Ace III
Bsp6 II
Bsa XI
BspKT5I
Eco57 I
HgiE II
BsmH I
Bca I
HinP1 I
Bms142 I
ClaI
Hha I
Hae II
SmiI
AcrI
AquI
AvaI
Eco88I
Nli3877 I
Bce83 I
Bca I
HinP1 I
Asp24 I
BsoJI
BarFI
ClaI
Cfr10 I
Hha I
NgoM I
SgrA I
Nae I
SauLPI
Nha I
Ace II
Bsp6 II
AgeI
BsrFI
Cfr10 I
PflMI
Uba1303 I
BsaO I
BspKT5I
Eco57 I
Gsp I
BsrI
MspA1 I
Pvu II
Mst I
EcoR124 II
BsoD I
Cfr14 I
Uba1303 I
Bli889I
Eae I
Eco52 I
XmaIII
Bsa0 I
GdI II
GdI II
Bst29 I
Bsu36 I
EcoR124 II
AcsI
Apo I
Bsg I
BspM I
Mst I
Cfr14 I
EcoD XXI
Msp20 I
Bli889I
Eae I
Bal I

Fig. 4 cont'd

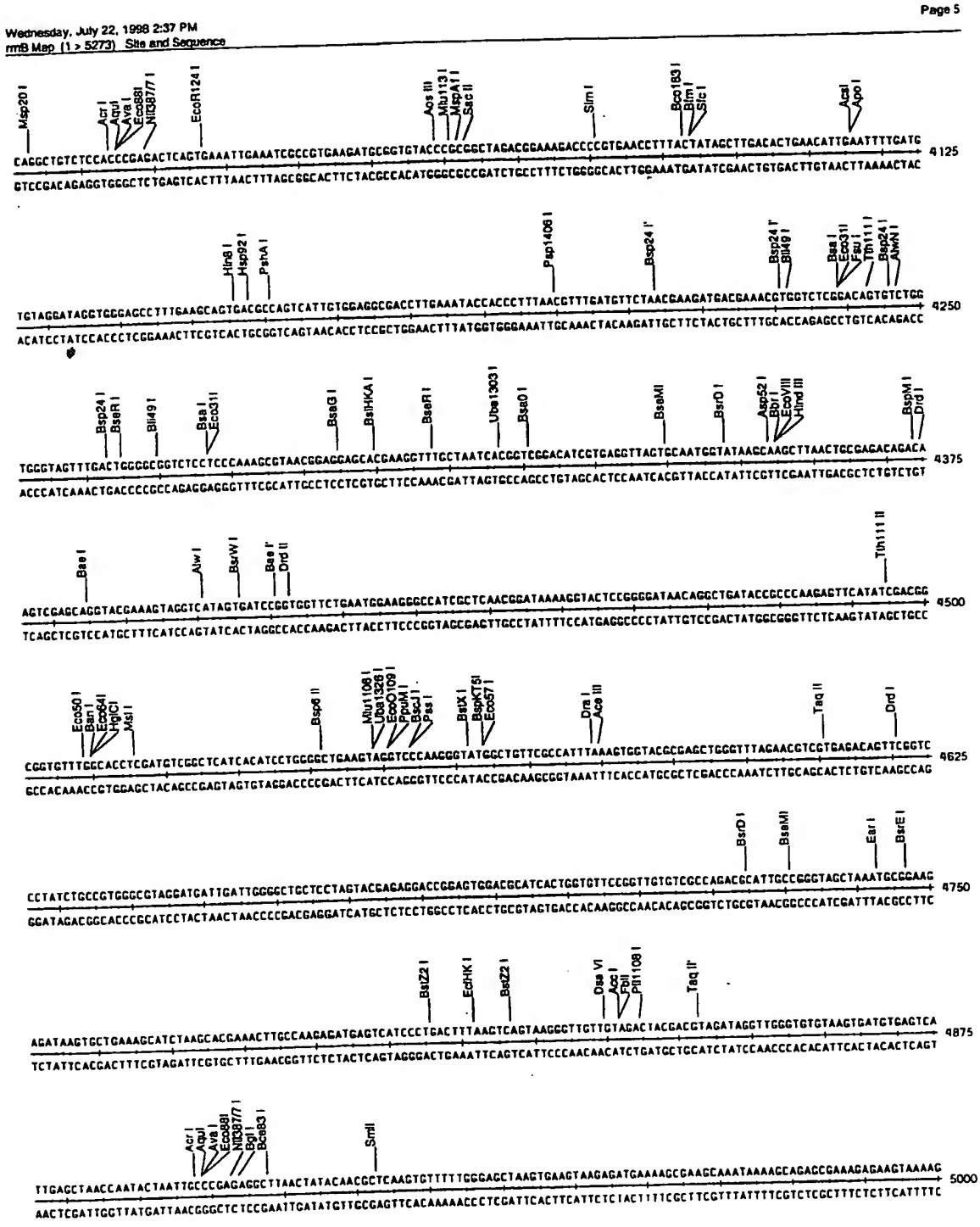

Wednesday, July 22, 1998 2:37 PM
rmB Map (1 > 5273) Site and Sequence
Page 5
Msp20 I
Acr I
Aqu I
Ava I
Eco88I
Nlll3877 I
EcoR124 I
Aos III
Mlu113 I
MspA1 I
Sac II
Sim I
Bco183 I
Blm I
Sfc I
AgeI
Apo I
4125
Hin8 I
Hsp92 I
PshA I
Psp1406 I
Bsp24 I'
Bsp24 I'
Bll49 I
Bsa I
Eco31I
Fau I
Tth111 I
Bsp24 I
AlwN I
4250
Bsp24 I
Bse8R I
Bll49 I
Bsa I
Eco31I
BsaG I
BsiHKA I
BseR I
Ube1303 I
Bse80 I
BseMI
BsrD I
Asp52 I
Bbr I
EcoVIII
Hind III
BspM I
Drd I
4375
Bse I
AlwI
BsrW I
Bse I'
Drd II
Tth111 II
4500
Eco50 I
Ban I
Eco64I
HgiCI
MsI I
Bsp6 II
Mlu1106 I
Uba1328 I
EcoO109 I
PpuM I
BscJ I
Pss I
BstX I
BspKT5I
Eco57 I
Dra I
Ace III
Taq II
Drd I
4625
BsrD I
BseMI
Esr I
BsrE I
4750
BstZ2 I
EcHK I
BstZ2 I
Dsa VI
Acc I
PstI
Pst1108 I
Taq II'
4875
Acr I
Aqu I
Ava I
Eco88I
Nlll3877 I
Bgl I
Bca83 I
SmII
5000

FIG. 4 Cont'd

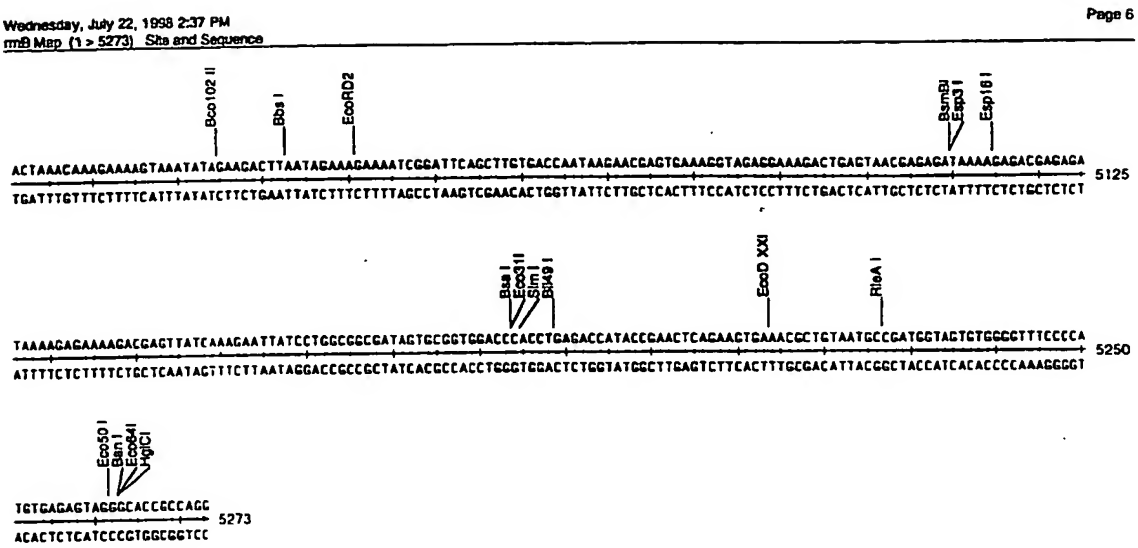

Wednesday, July 22, 1998 2:37 PM
Page 6
rm8 Map (1 > 5273) Sits and Sequence
Bco102 II
Bbs I
EcoRD2
BsmBI
Esp3 I
Esp16 I
ACTAAACAAAGAAAAGTAAATATAGAAGACTTAATAGAAAGAAAATCGGATTCAGCTTGTGACCAATAAGAACGAGTCAAAGGTAGACGAAAGACTGAGTAACGACAGATAAAAGACACGAGACA
5125
TGATTTGTTTCTTTTCATTTATATCTTCTGAATTATCTTTCTTTTAGCCTAAGTCGAACACTGGTTATTCTTGCTCACTTTCCATCTCCTTTCTGACTCATTGCTCTCTATTTTCTCTGCTCTCT
Bsa I
Eco31I
Sim I
BsrG I
EcoO XXI
RleA I
TAAAAGAGAAAAGACGAGTTATCAAAGAATTATCCTGGCGGCGATAGTGCGGTGGACCCACCTGAGACCATACCGAACTCAGAAGTGAAACGCCTGTAATGCCGATGGTAGTGTGGGGTTTCCCCA
5250
ATTTTCTCTTTTCTGCTCAATAGTTTCTTAATAGGACCGCCGCTATCACGCCACCTGGGTGGACTCTGGTATGGCTTGAGTCTTCACTTTGCGACATTACGGCTACCATCACACCCCAAAGGGGT
Eco50 I
Ban I
Eco64I
HgiCI
TGTGAGAGTAGGGCACCGCCAGG
5273
ACACTCTCATCCCGTGGCGGTCC

FIG. 5

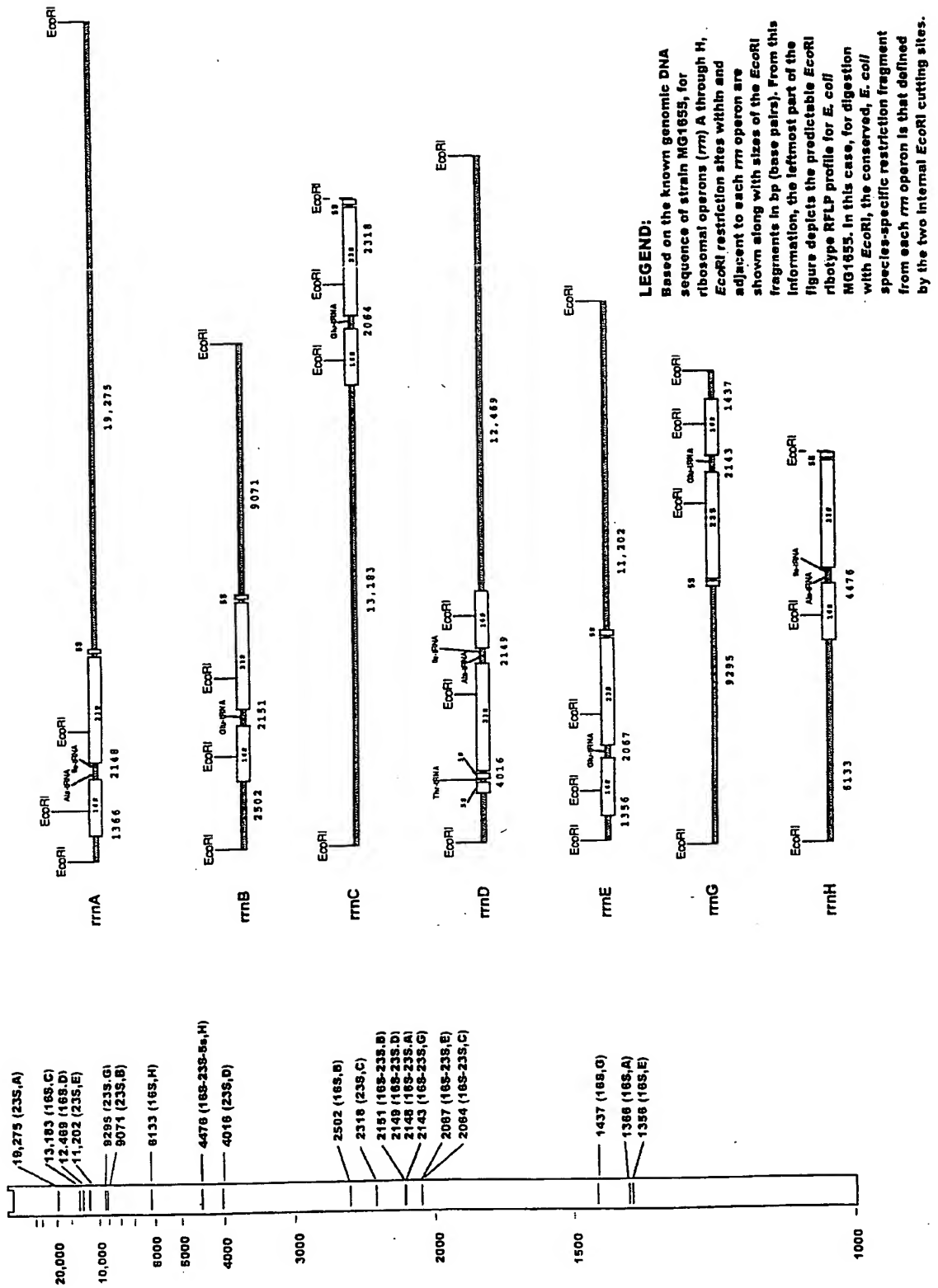
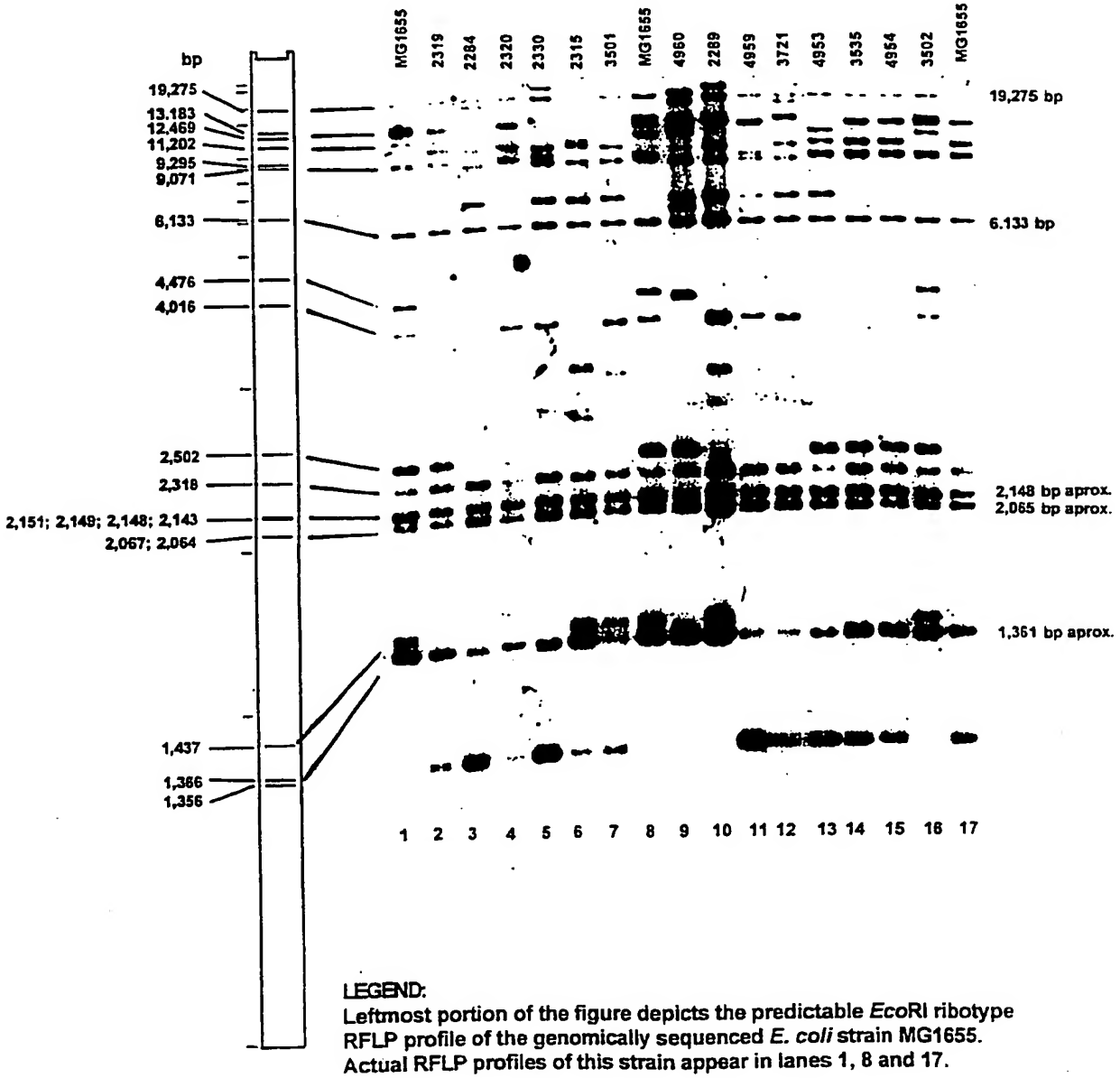

The 7 ribosomal RNA operons of the genomically sequenced E. coli strain MG1655
19,275 (23S,A)
13,183 (16S,C)
12,469 (16S,D)
11,202 (23S,E)
9295 (23S,G)
9071 (23S,B)
8133 (16S,H)
4476 (16S-23S-5s,H)
4016 (23S,D)
2502 (16S,B)
2318 (23S,C)
2151 (16S-23S,B)
2148 (16S-23S,D)
2148 (16S-23S,A)
2143 (16S-23S,G)
2067 (16S-23S,E)
2064 (16S-23S,C)
1437 (16S,G)
1366 (16S,A)
1358 (16S,E)
20,000
10,000
8000
5000
4000
3000
2000
1500
1000
rrnA
rrnB
rrnC
rrnD
rrnE
rrnG
rrnH
EcoRI
LEGEND:
Based on the known genomic DNA sequence of strain MG1655, for ribosomal operons (rrn) A through H, EcoRI restriction sites within and adjacent to each rrn operon are shown along with sizes of the EcoRI fragments in bp (base pairs). From this information, the leftmost part of the figure depicts the predictable EcoRI ribotype RFLP profile for E. coli MG1655. In this case, for digestion with EcoRI, the conserved, E. coli species-specific restriction fragment from each rrn operon is that defined by the two internal EcoRI cutting sites.

FIG. 6

EcoRI ribotype RFLPs of *E. coli* isolates from diverse sources, including the genomically sequenced strain MG1655.



LEGEND:
Leftmost portion of the figure depicts the predictable *Eco*RI ribotype RFLP profile of the genomically sequenced *E. coli* strain MG1655. Actual RFLP profiles of this strain appear in lanes 1, 8 and 17.

Fig. 7

DNA sequence and restriction map of the *E. coli* MG1655 *rrnA* operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less.

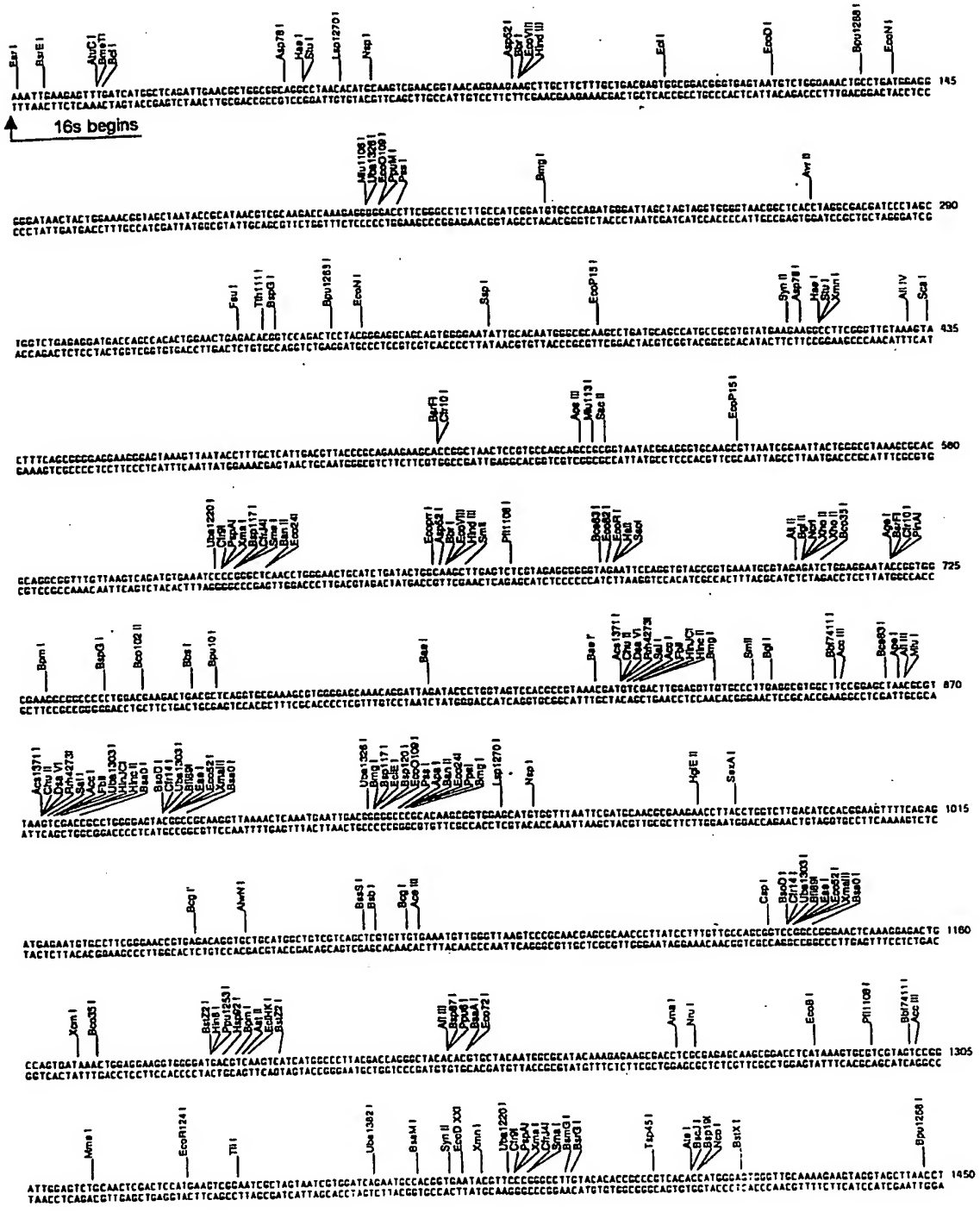


FIG. 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnA* operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less (continued 1).

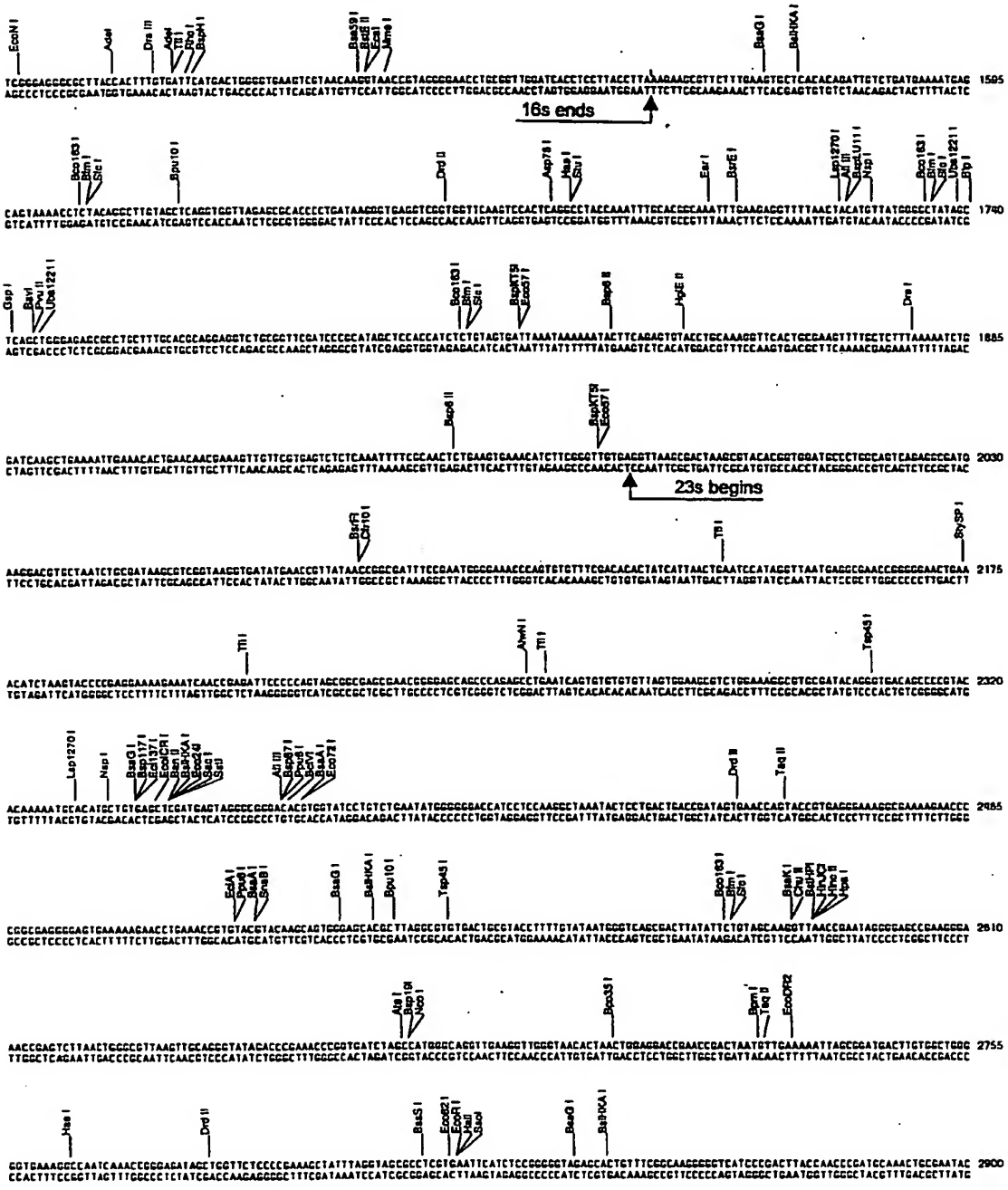


Fig. 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnA* operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less (continued 2).

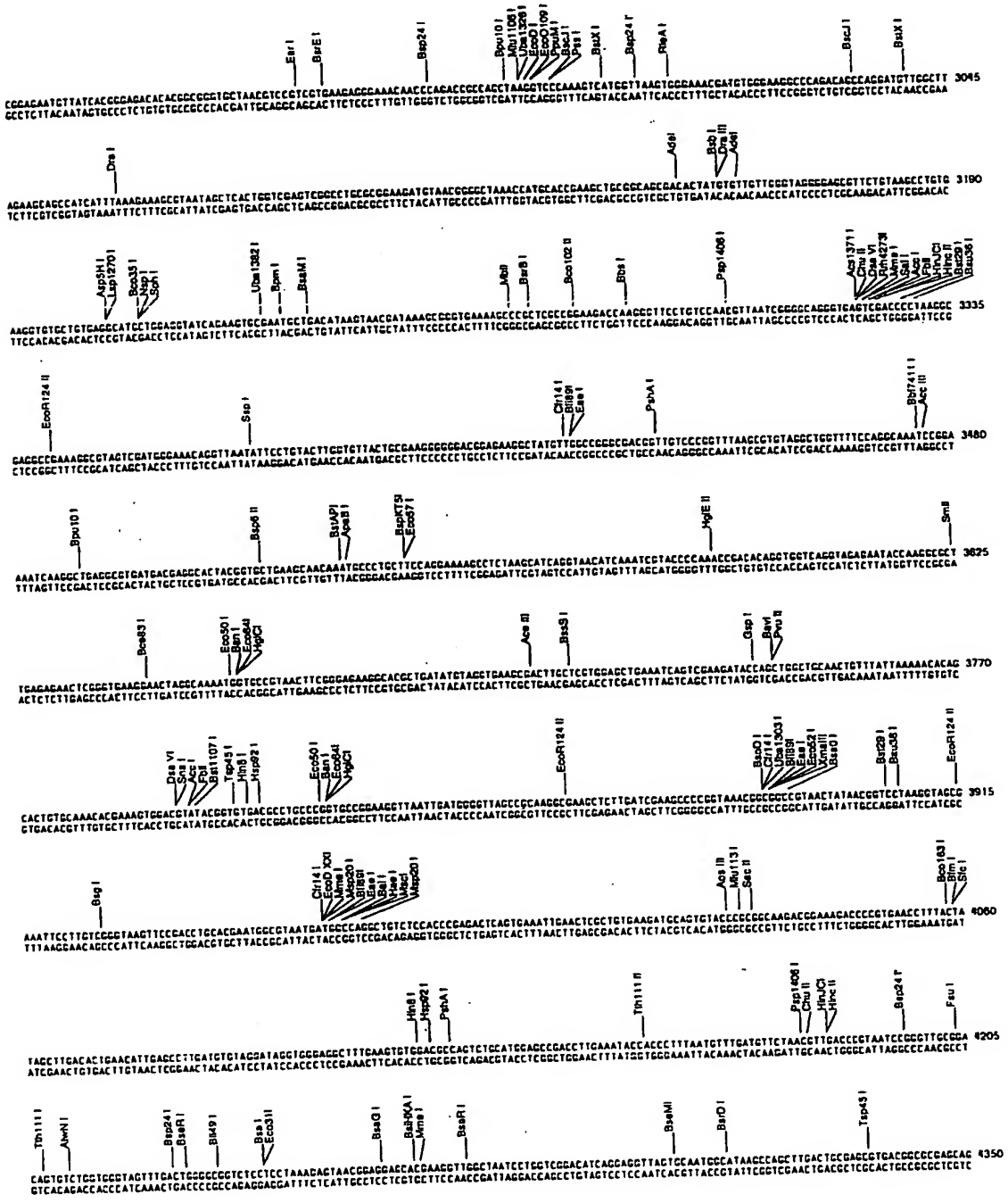


Fig. 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnA* operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less (continued 3).

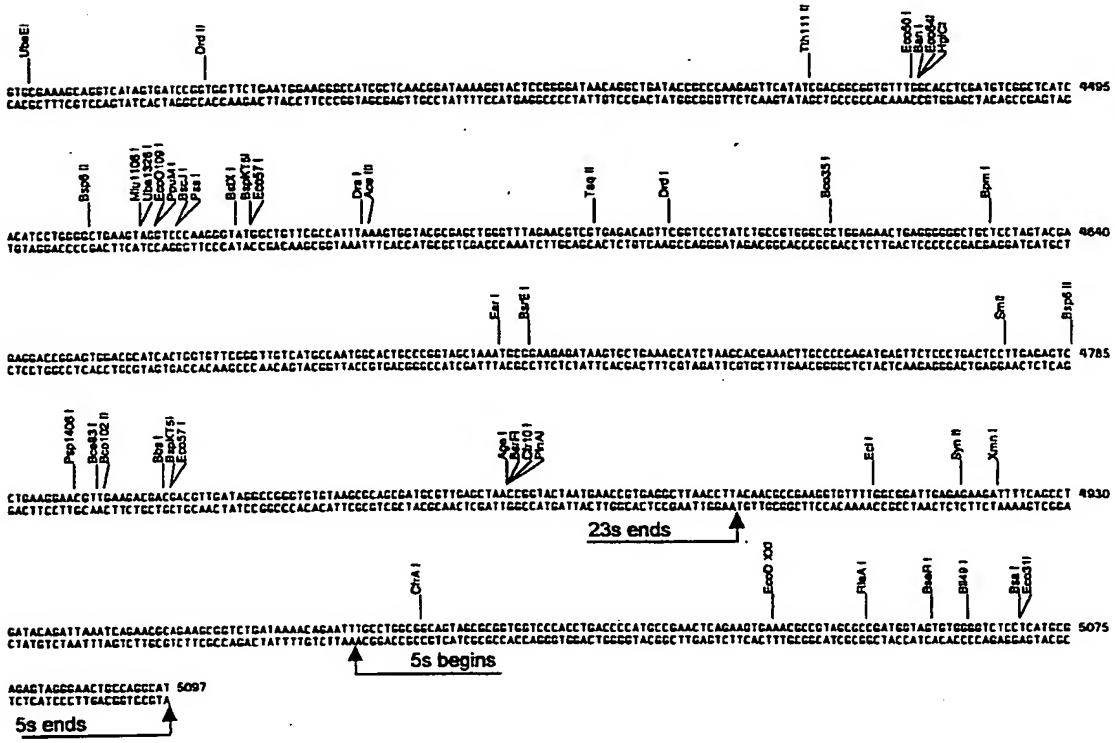


Fig. 7 cont'd

Restriction map of the *rrnA* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less.

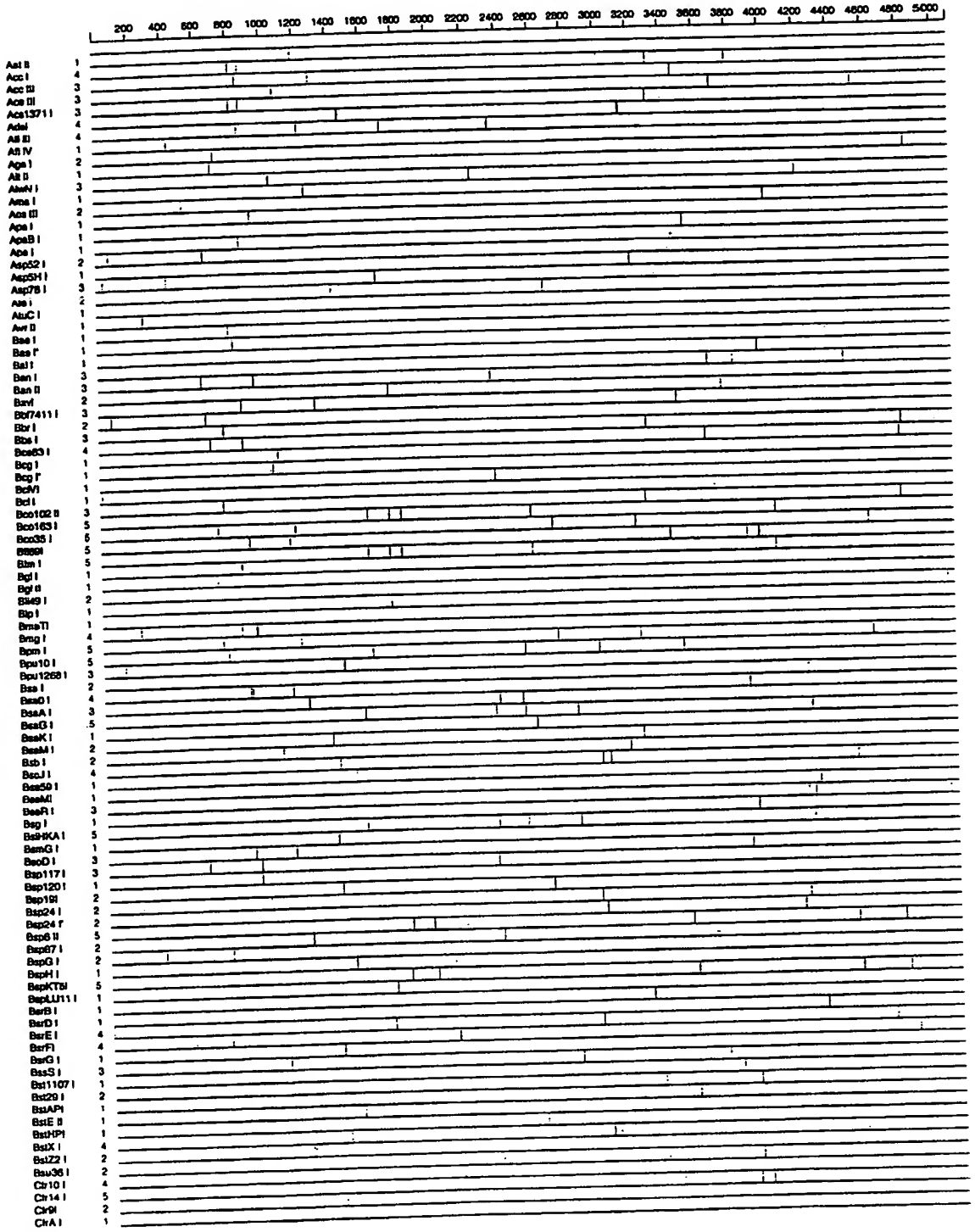


Fig. 7 cont'd

Restriction map of the *rrnA* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less (continued 1).

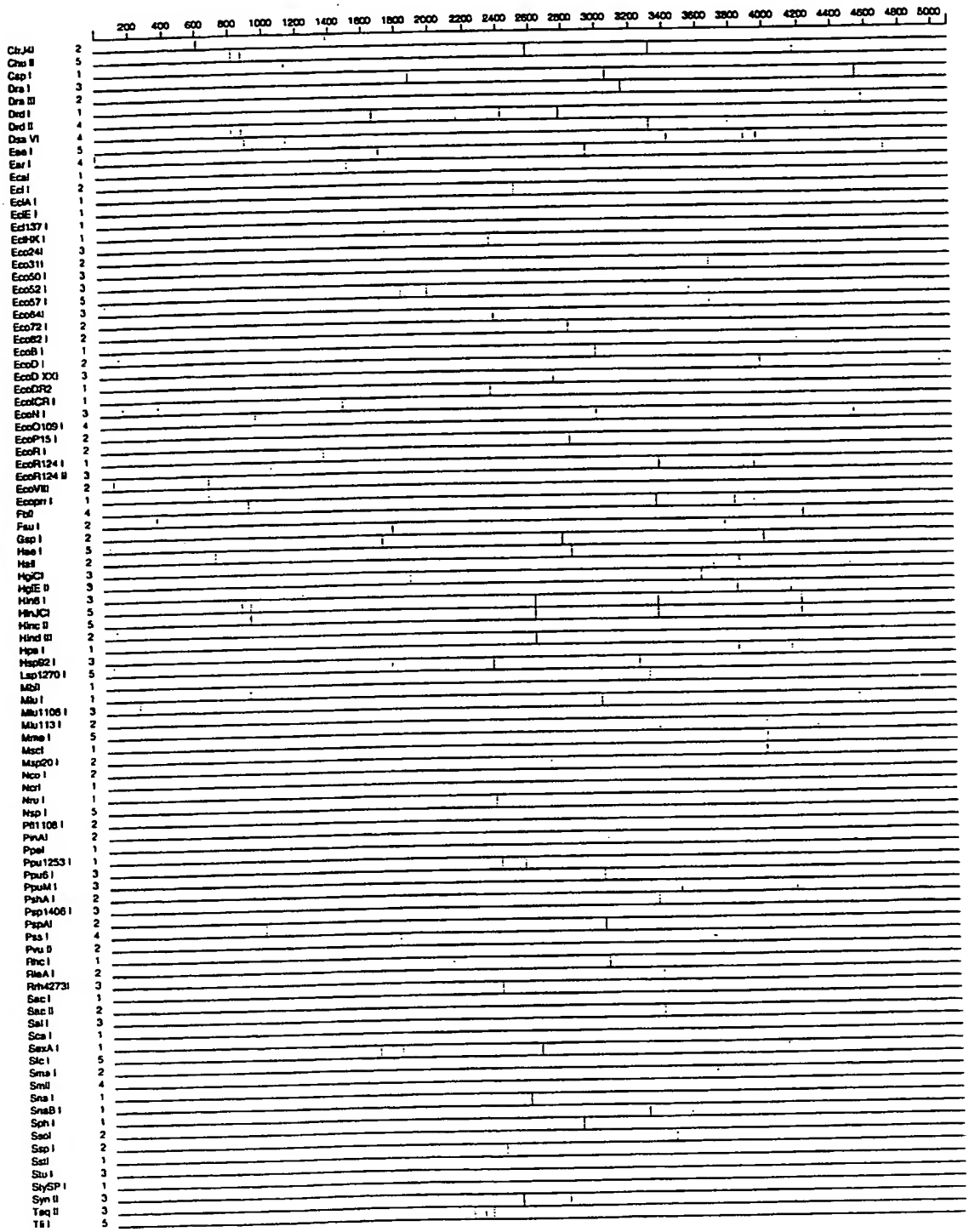


FIG. 7 Cont'd

Restriction map of the *rrnA* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less (continued 2).

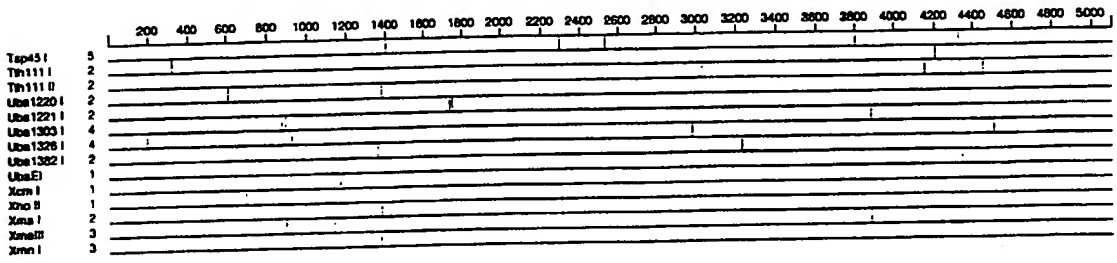


Fig. 7 Cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnB* operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less.

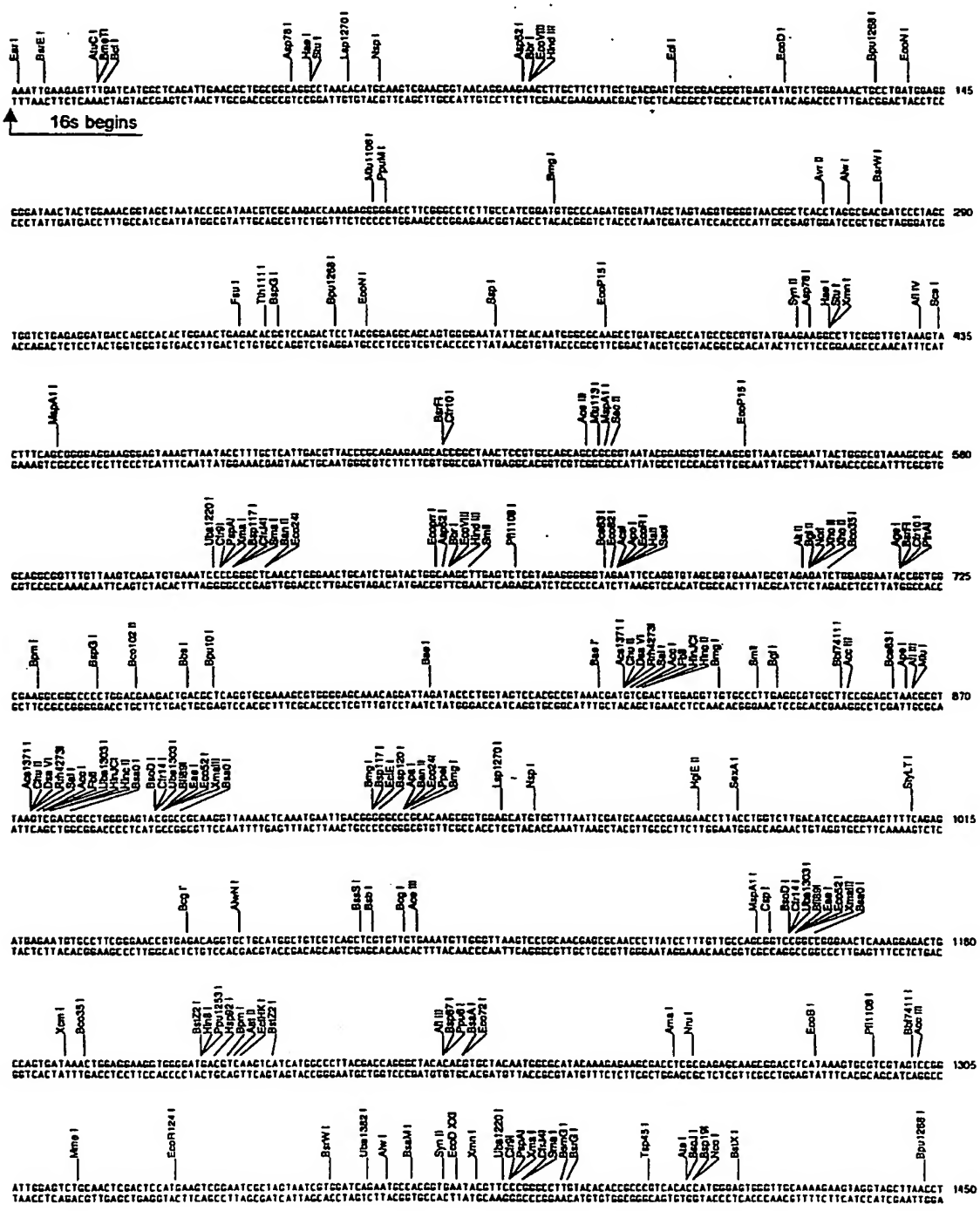


Fig. 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnB* operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less (continued 1).

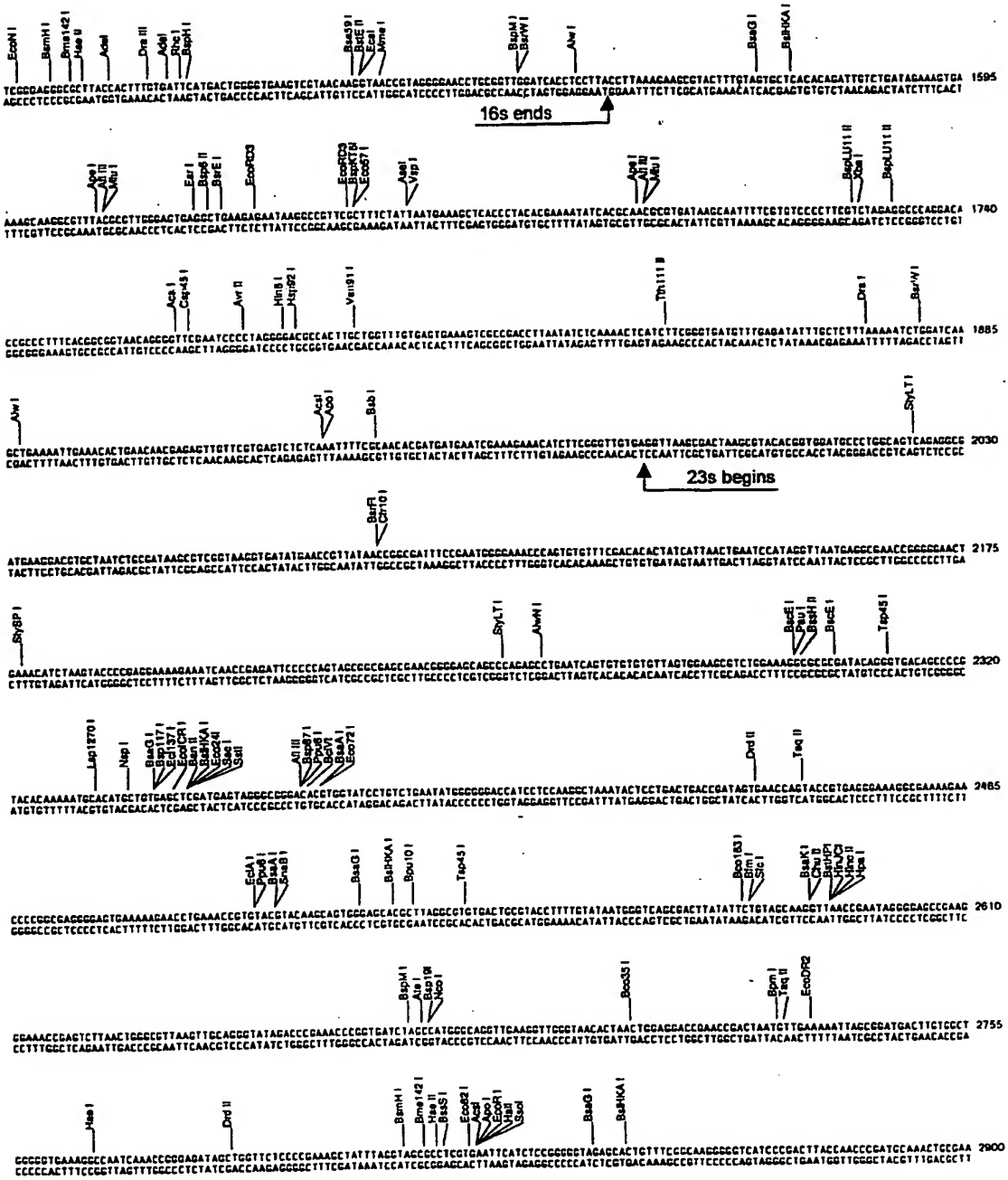


Fig. 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnB* operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less (continued 2).

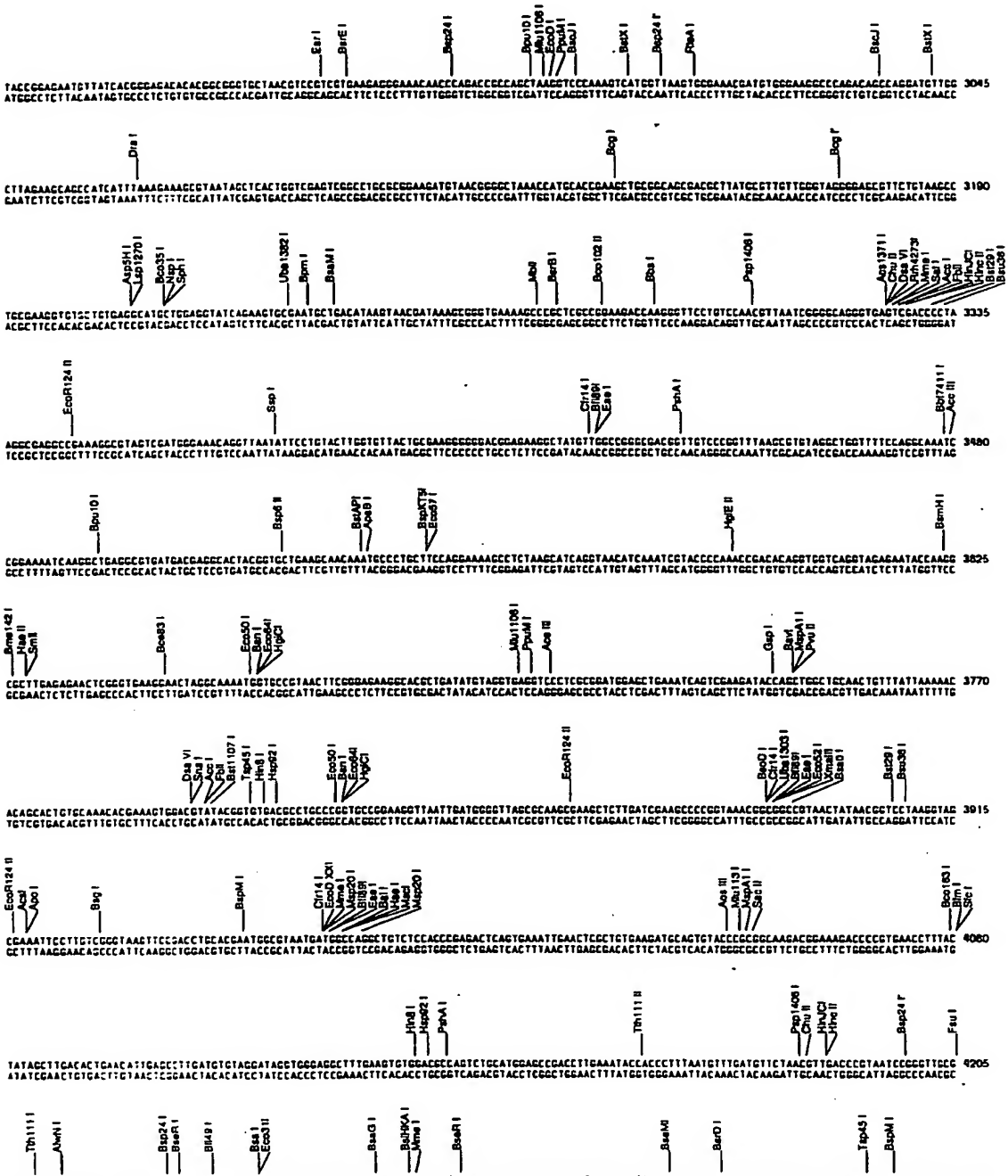


Fig. 7 Cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnB* operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less (continued 3).

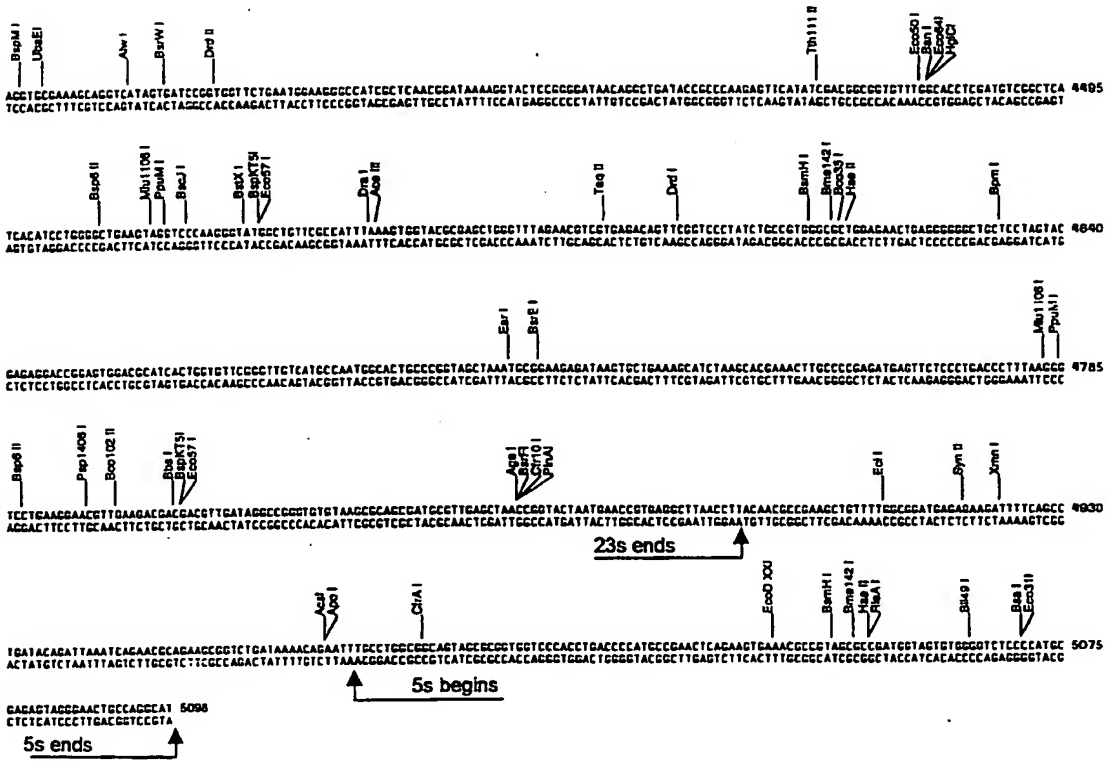


Fig. 7 cont'd

Restriction map of the *rrnB* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less.

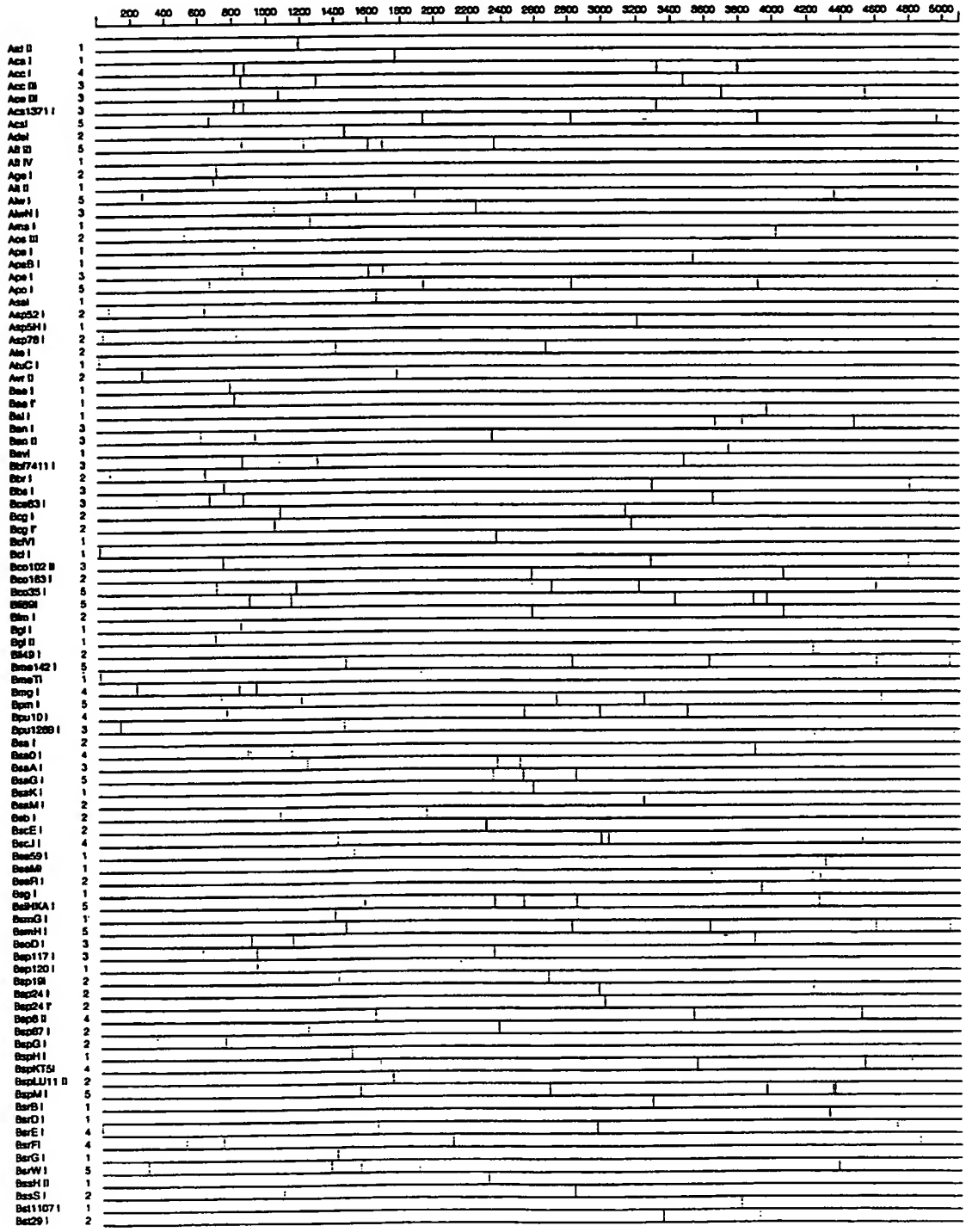


Fig. 7 cont'd

Restriction map of the *rrnB* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less (continued 1).

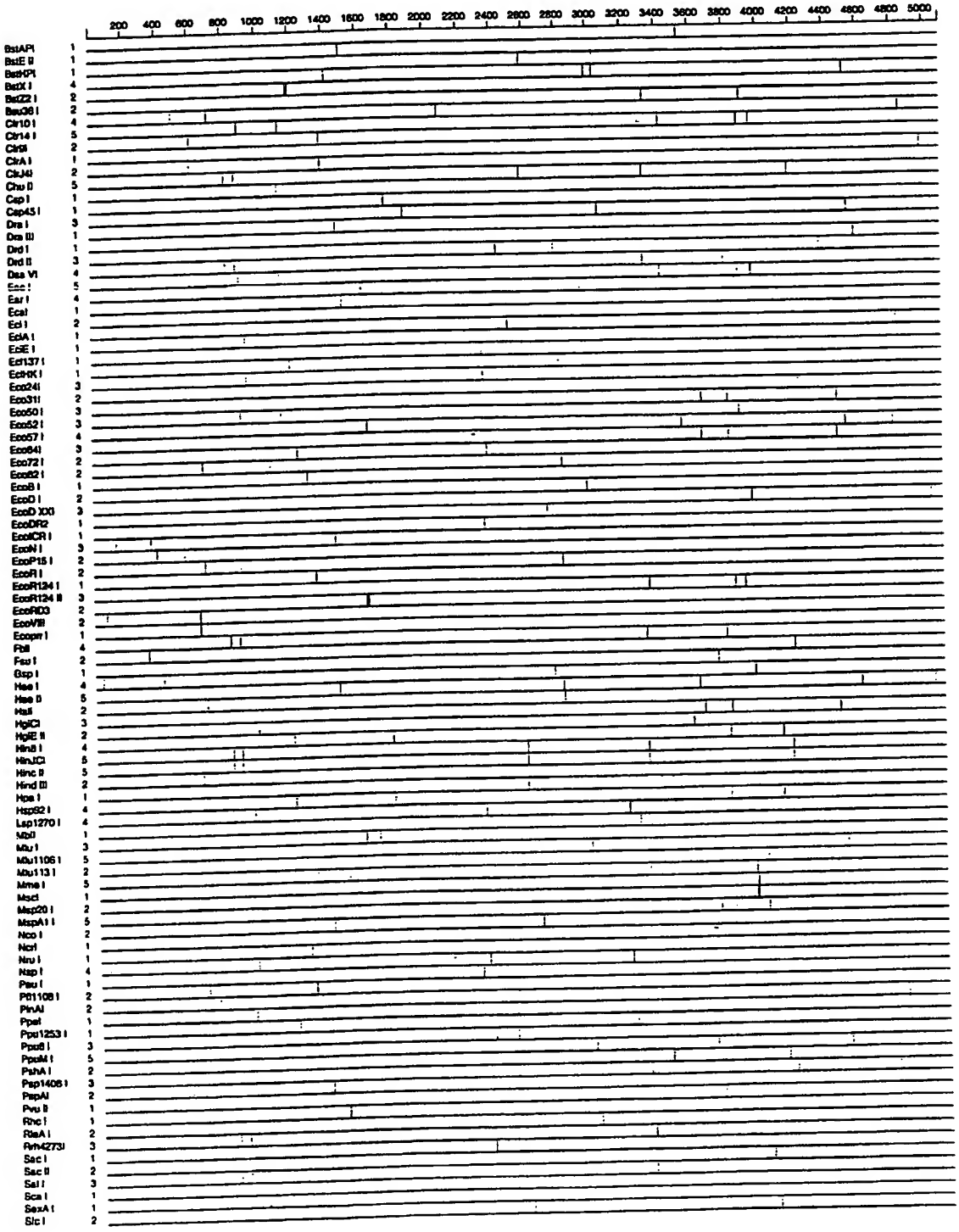


Fig. 7 cont'd

Restriction map of the *rrnB* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less (continued 2).

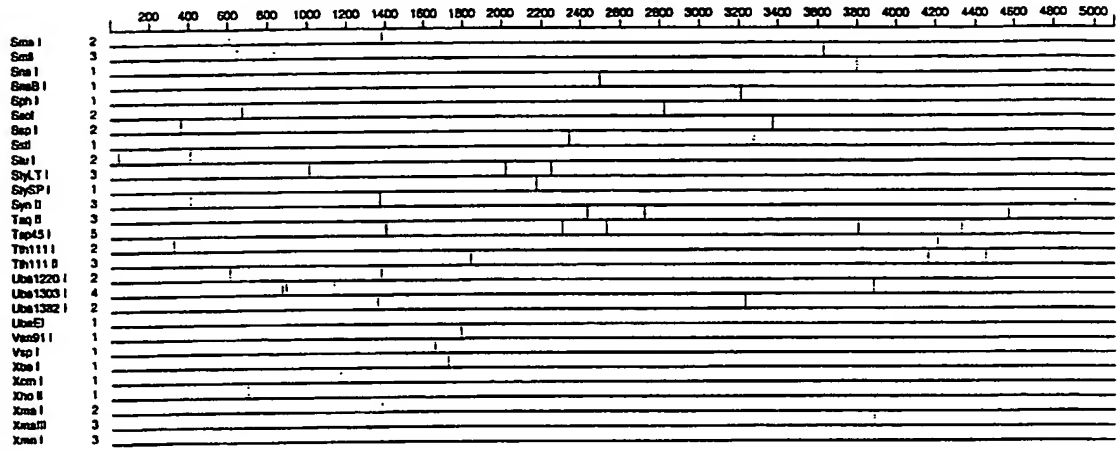


Fig. 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnC* operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less.

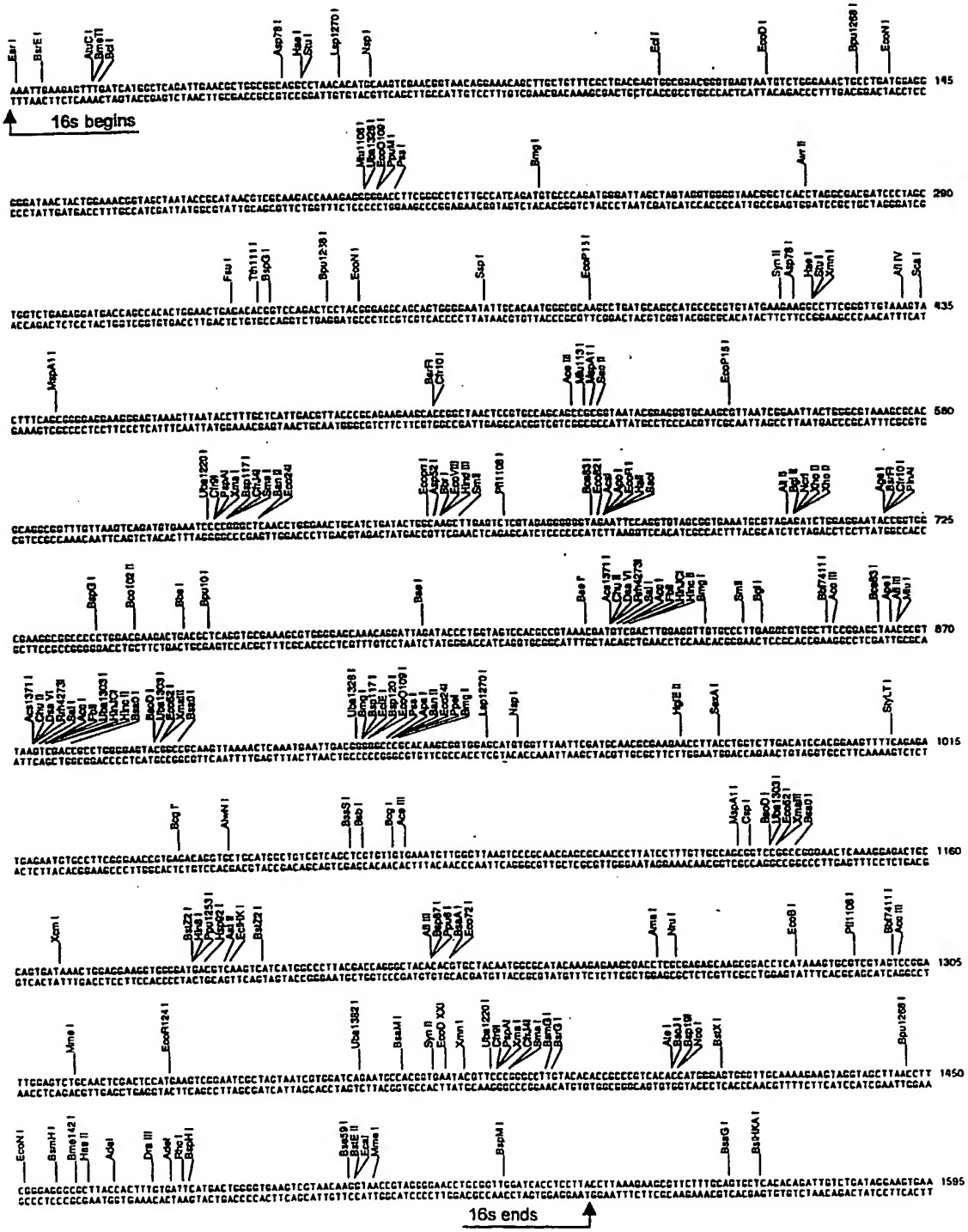


Fig. 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnC* operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less (continued 1).

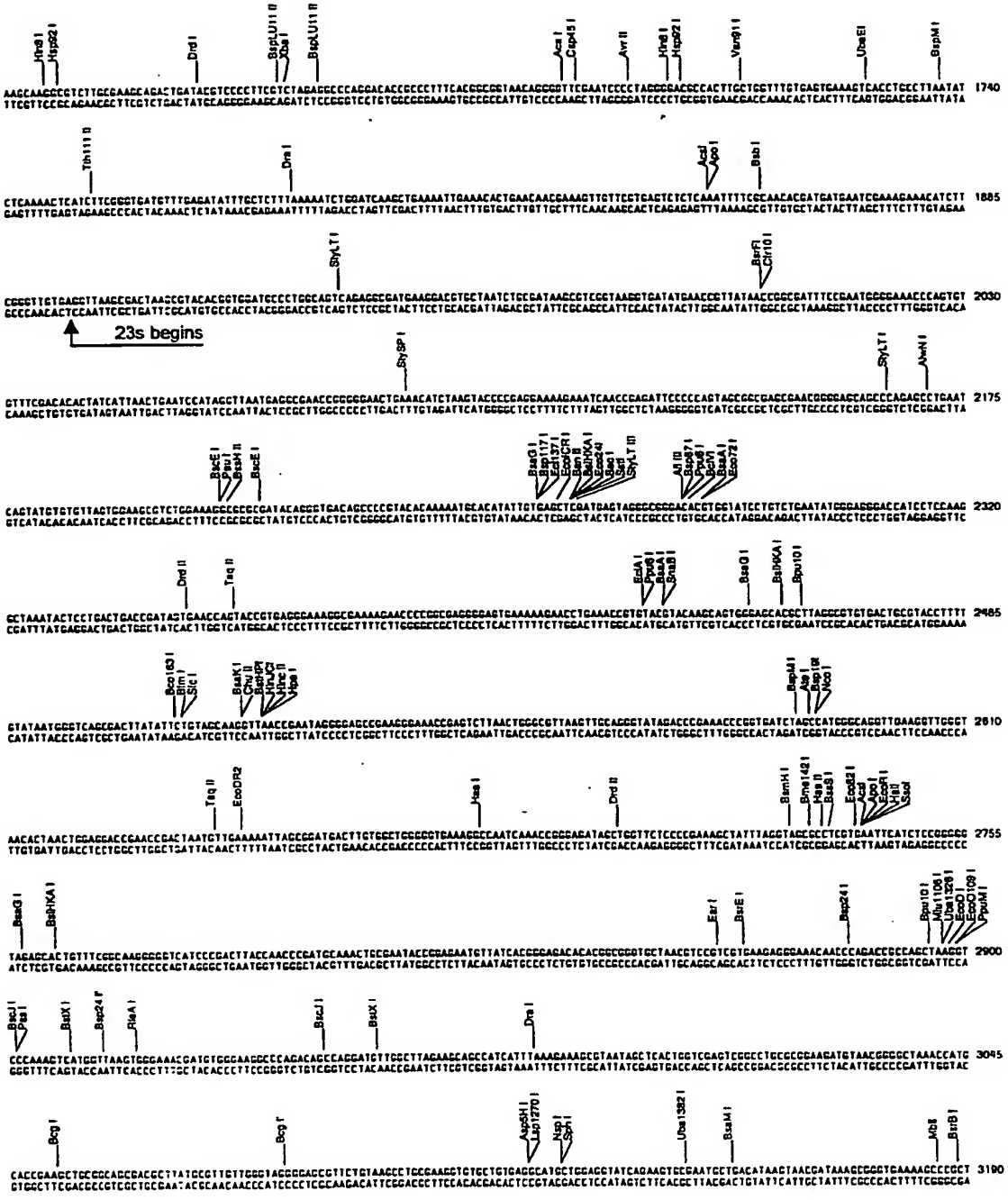


Fig. 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnC* operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less (continued 2).

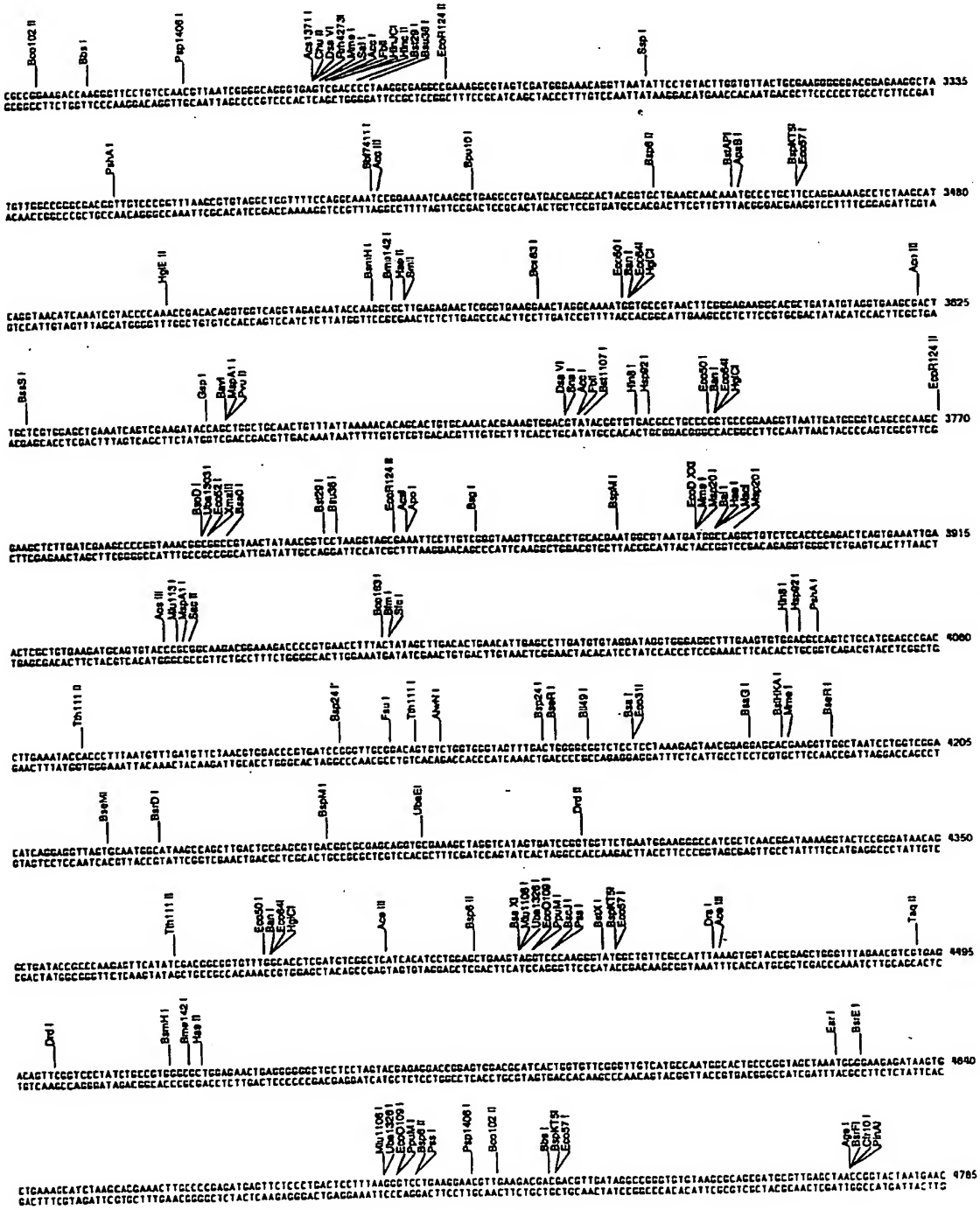


Fig. 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnC* operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less (continued 3).

Fig. 7 cont'd

Restriction map of the *rrnC* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less.

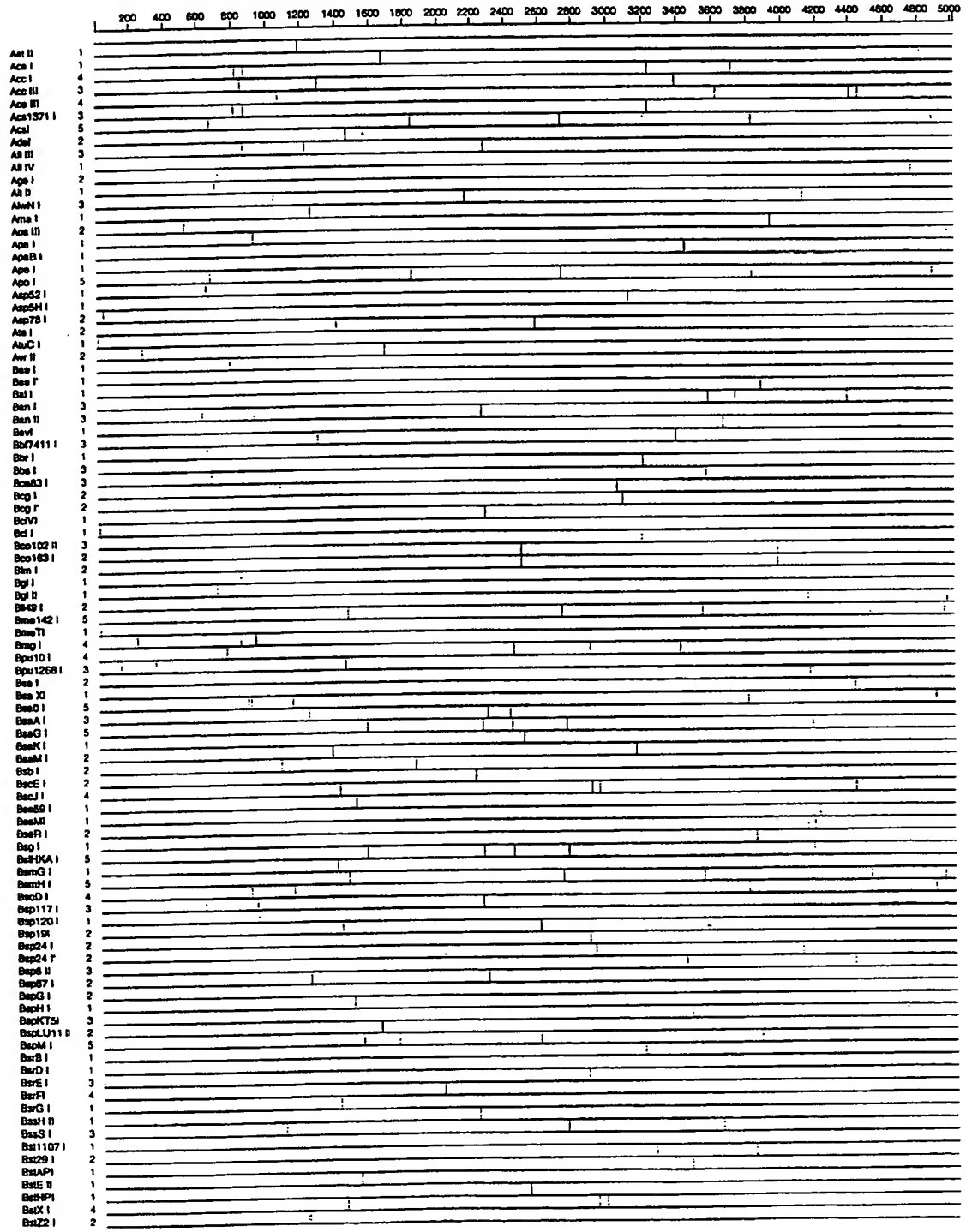


Fig. 7 cont'd

Restriction map of the *rrnC* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less (continued 1).

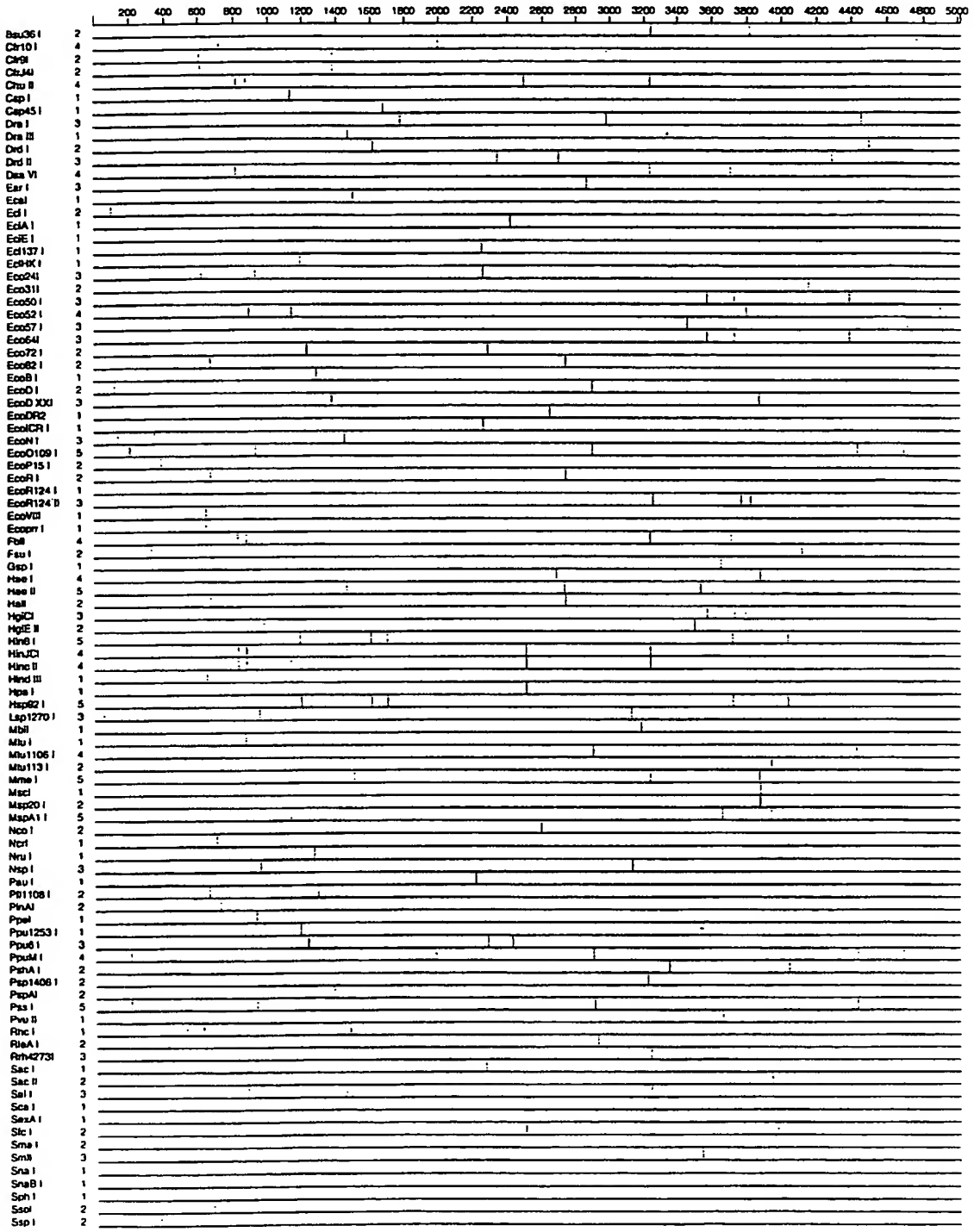


Fig. 7 cont'd

Restriction map of the *rrnC* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less (continued 2).

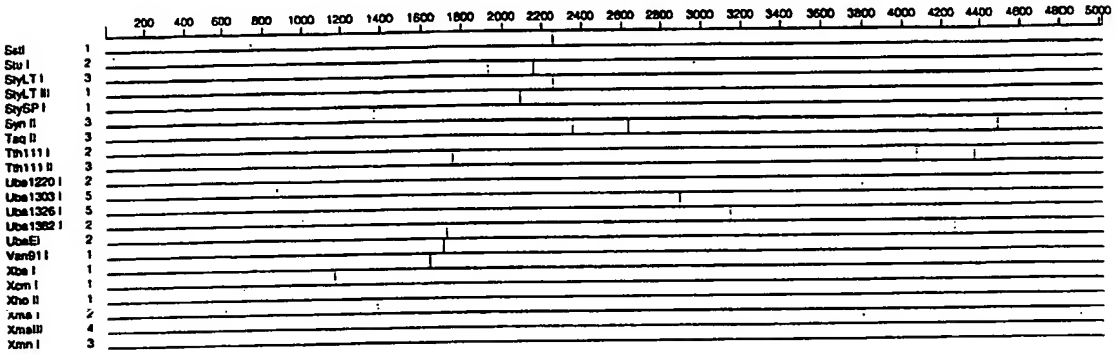


Fig. 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rmD* operon (5s-spacer-23s-spacer-16s), with restriction sites noted for enzymes cutting 5 times or less.

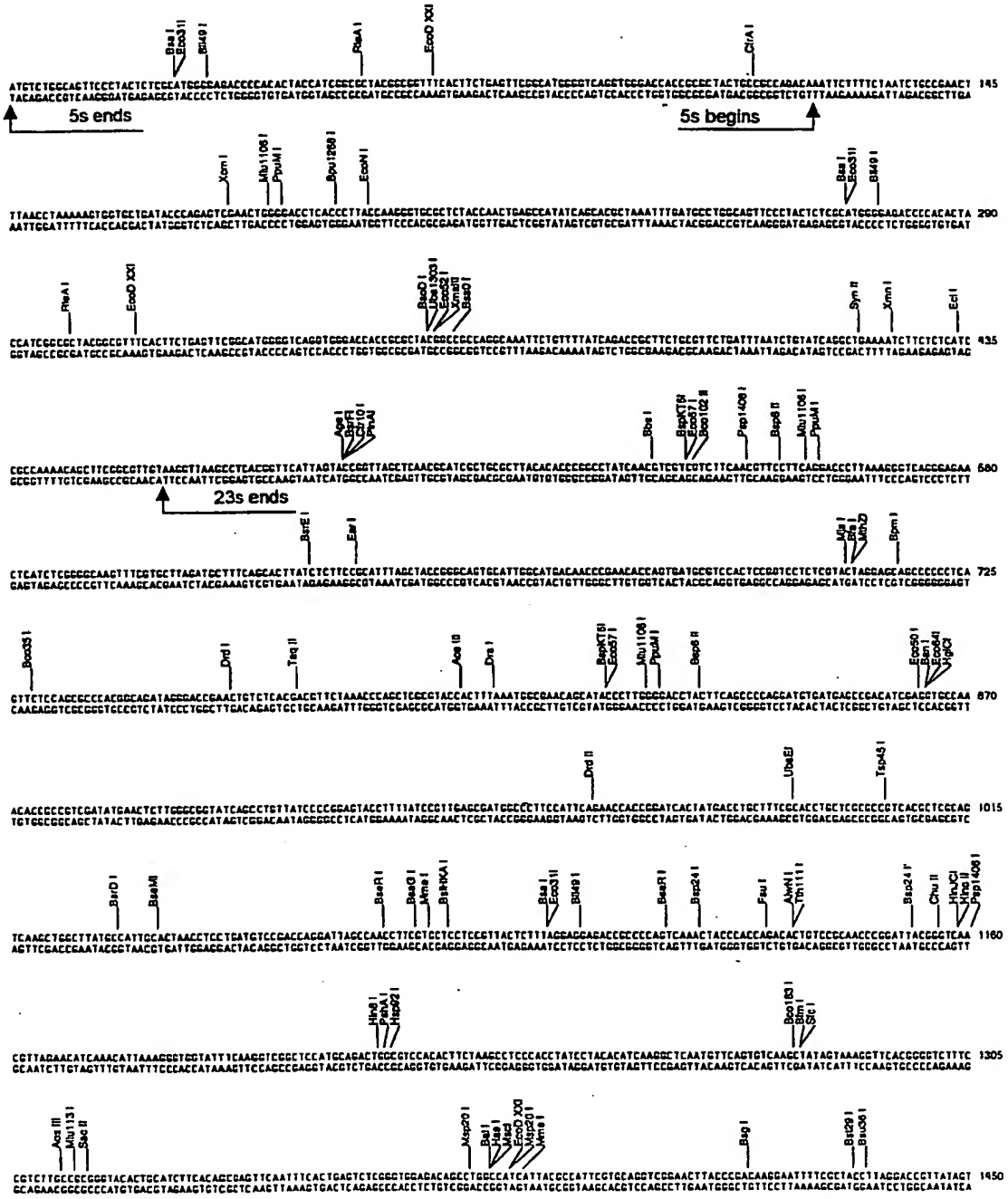


Fig. 7 contd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnD* operon (5s-spacer-23s-spacer-16s), with restriction sites noted for enzymes cutting 5 times or less (continued 1).

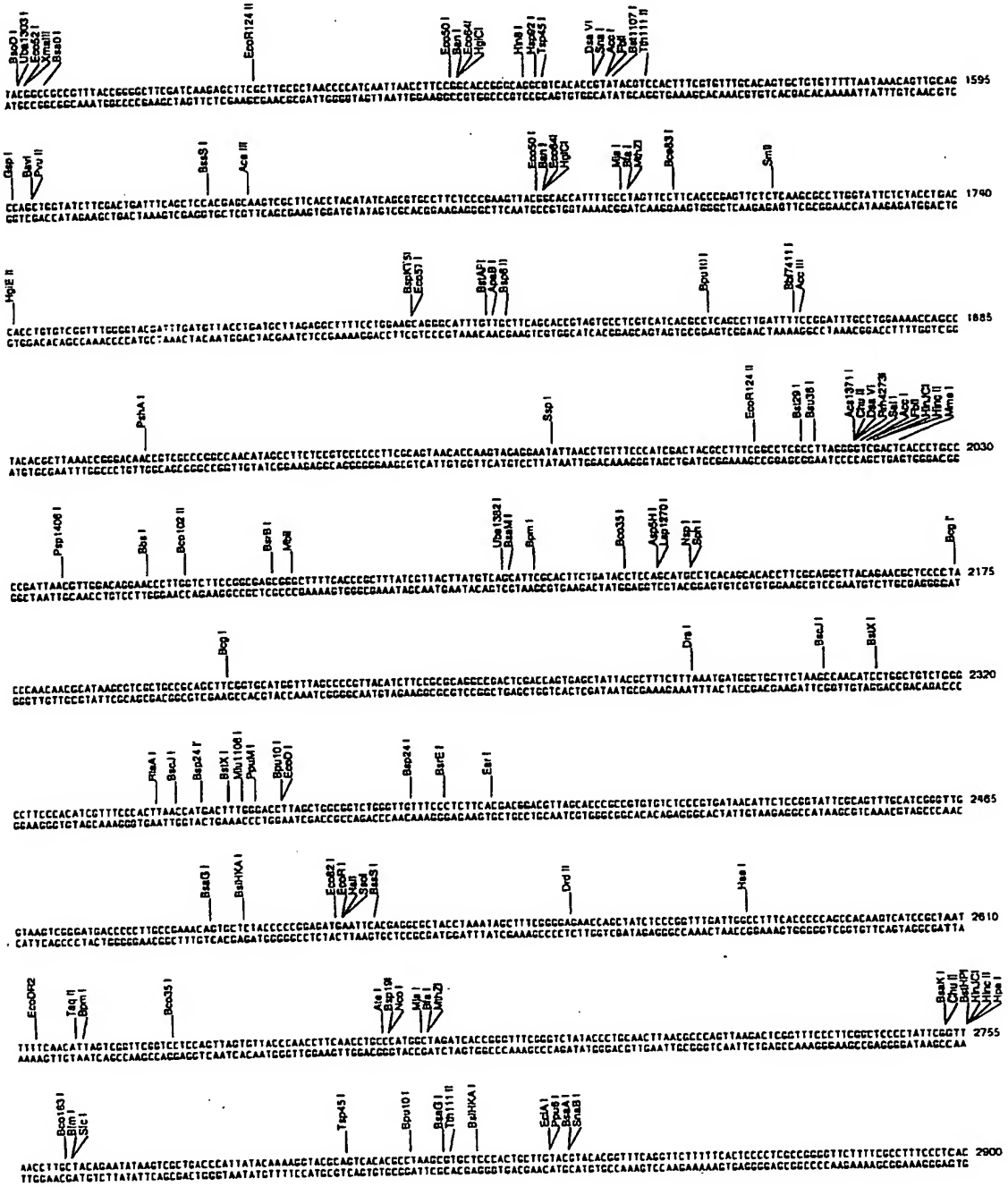


Fig. 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnD* operon (5s-spacer-23s-spacer-16s), with restriction sites noted for enzymes cutting 5 times or less (continued 2).

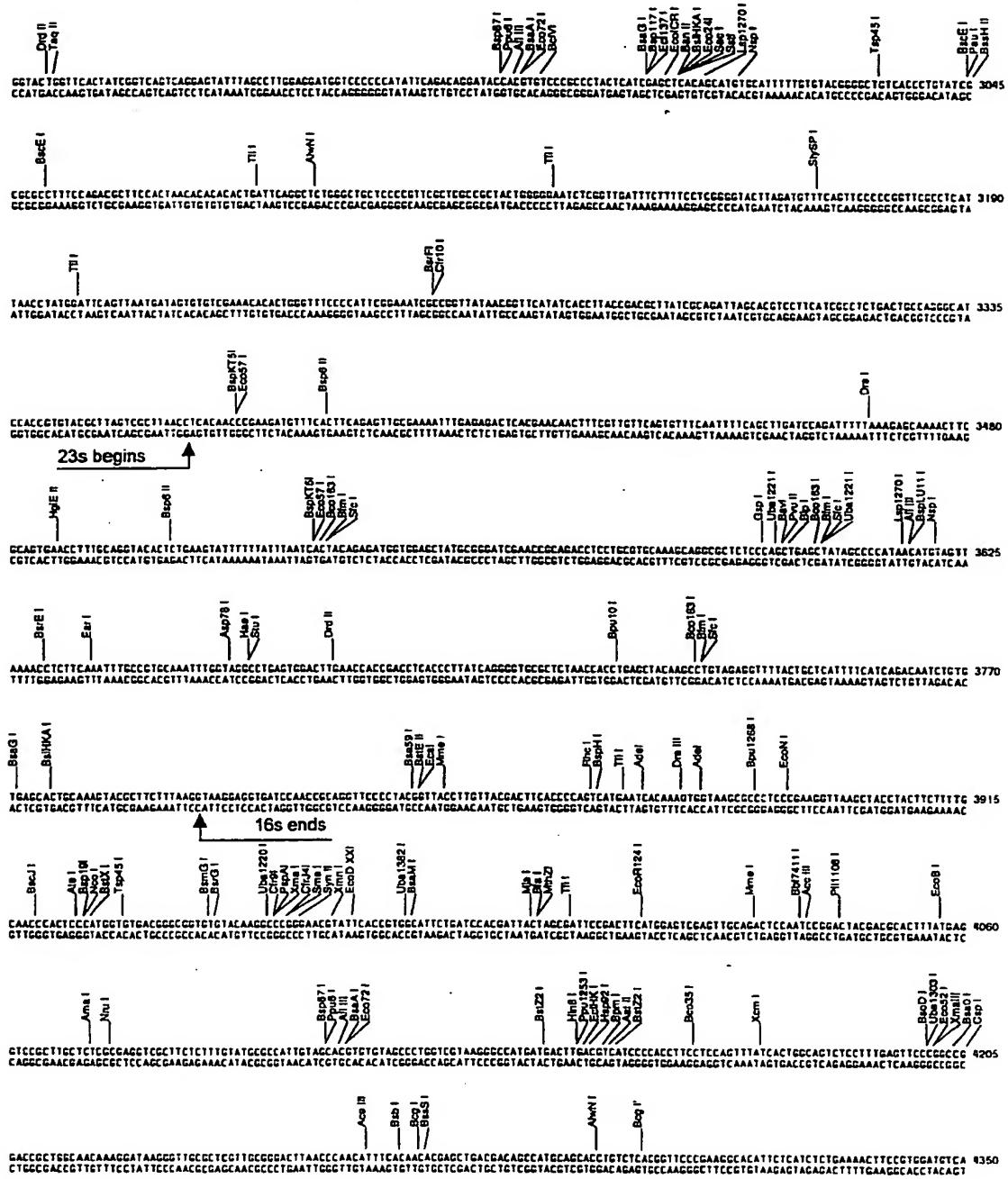


Fig. 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnD* operon (5s-spacer-23s-spacer-16s), with restriction sites noted for enzymes cutting 5 times or less (continued 3).

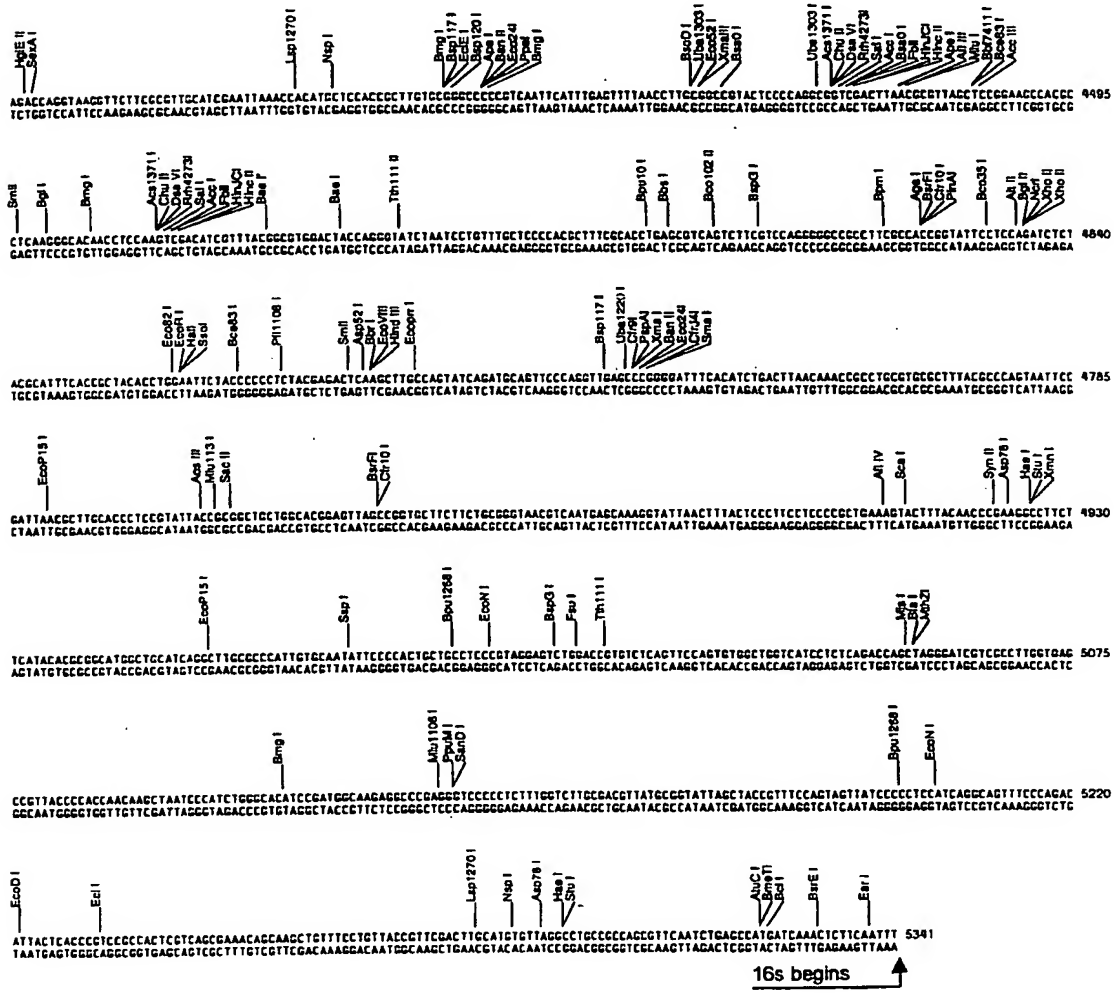


Fig. 7 cont'd

Restriction map of the *rrnD* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less.

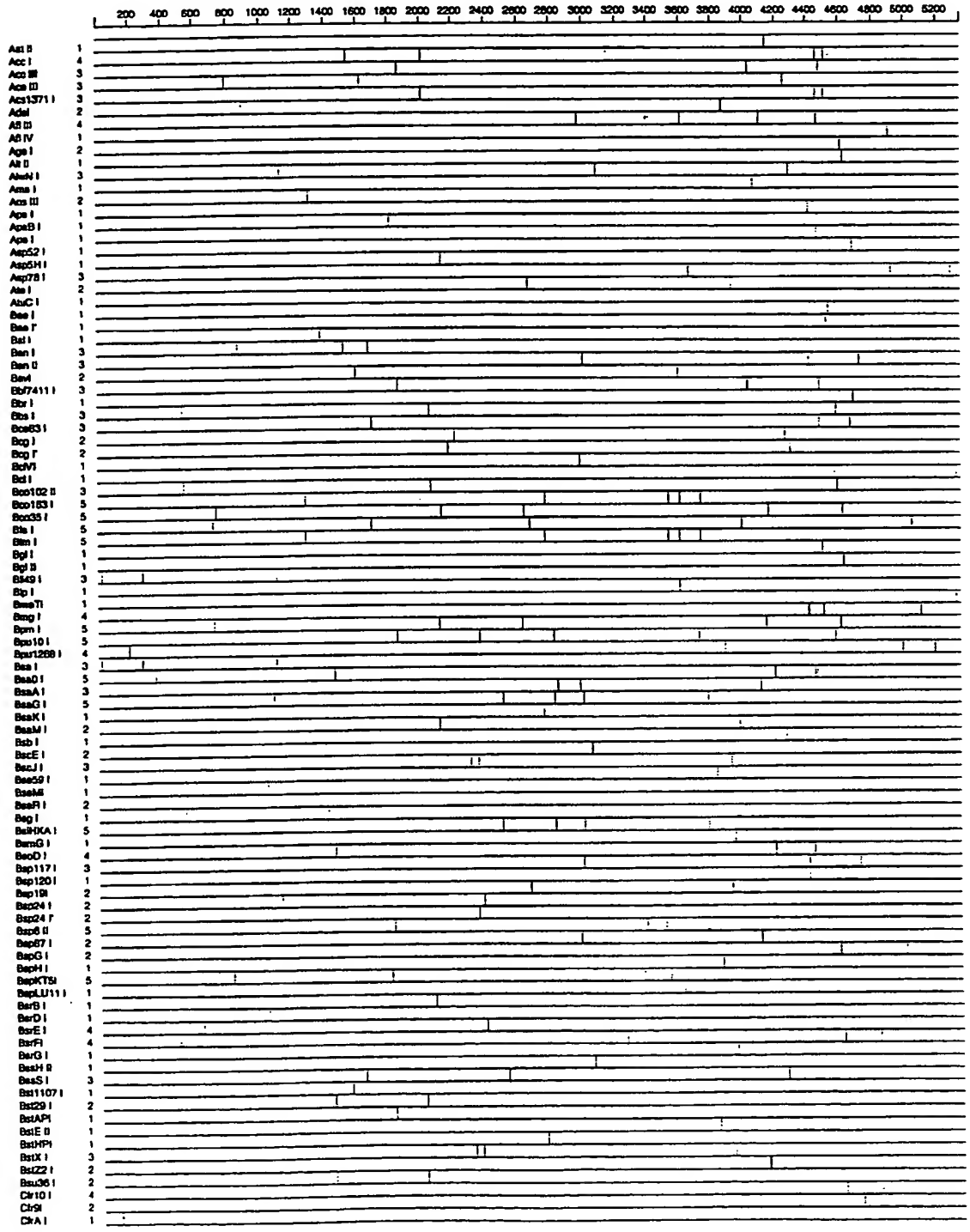


Fig. 7 cont'd

Restriction map of the *rrnD* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less (continued 1).

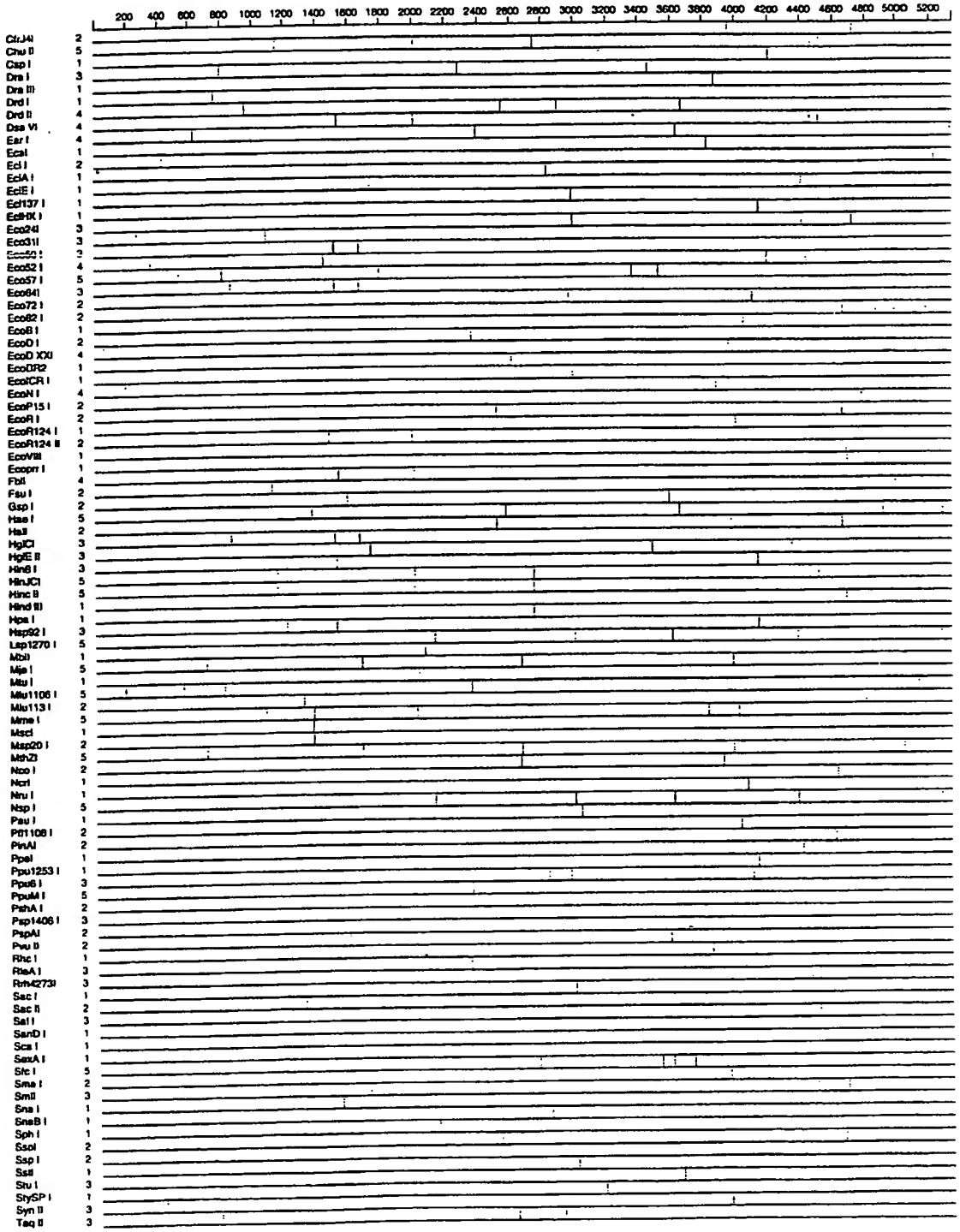


Fig. 7 cont'd

Restriction map of the *rrnD* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less (continued 2).

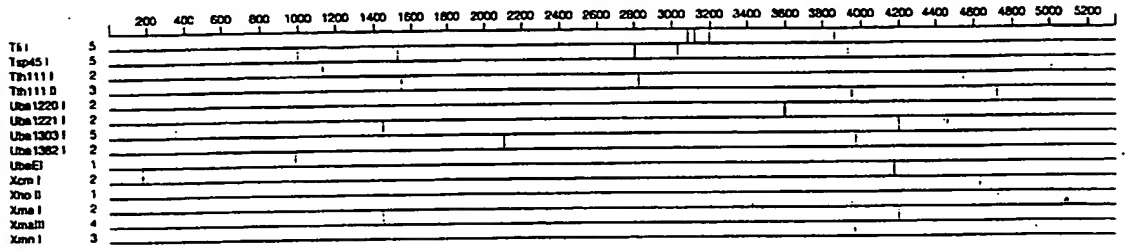


Fig. 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnE* operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less.

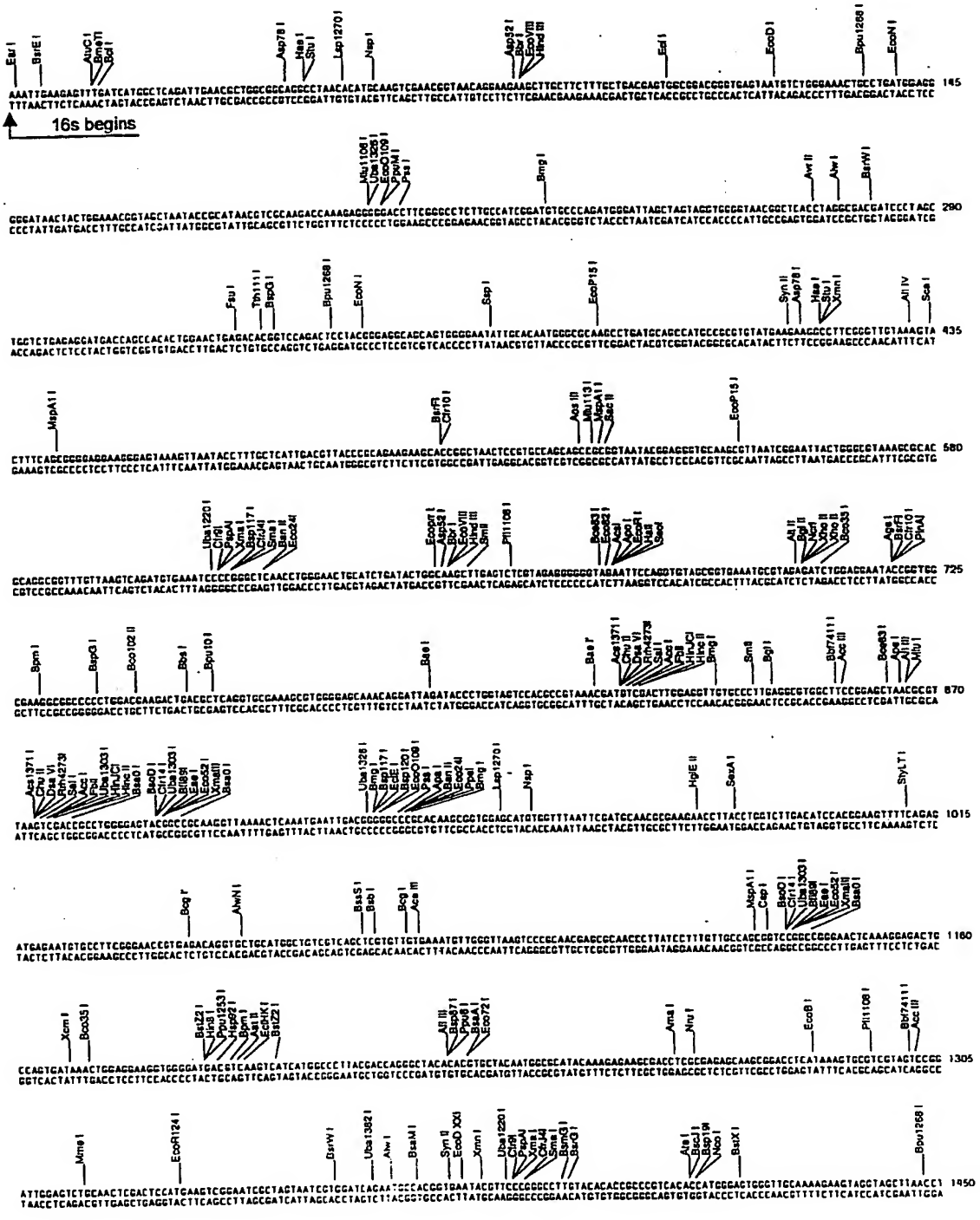


Fig. 7 Cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnE* operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less (continued 1).

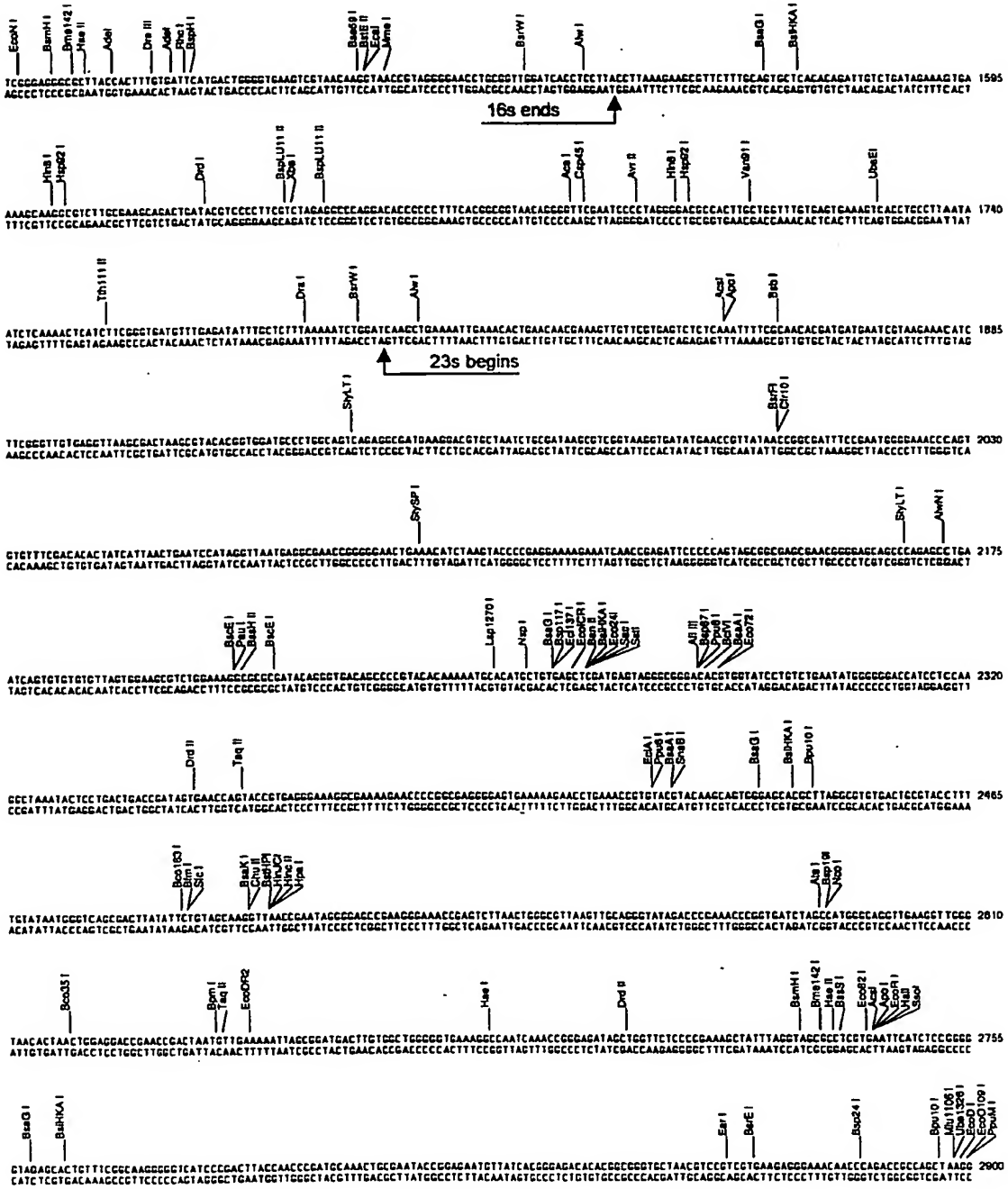


Fig. 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnE* operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less (continued 2).

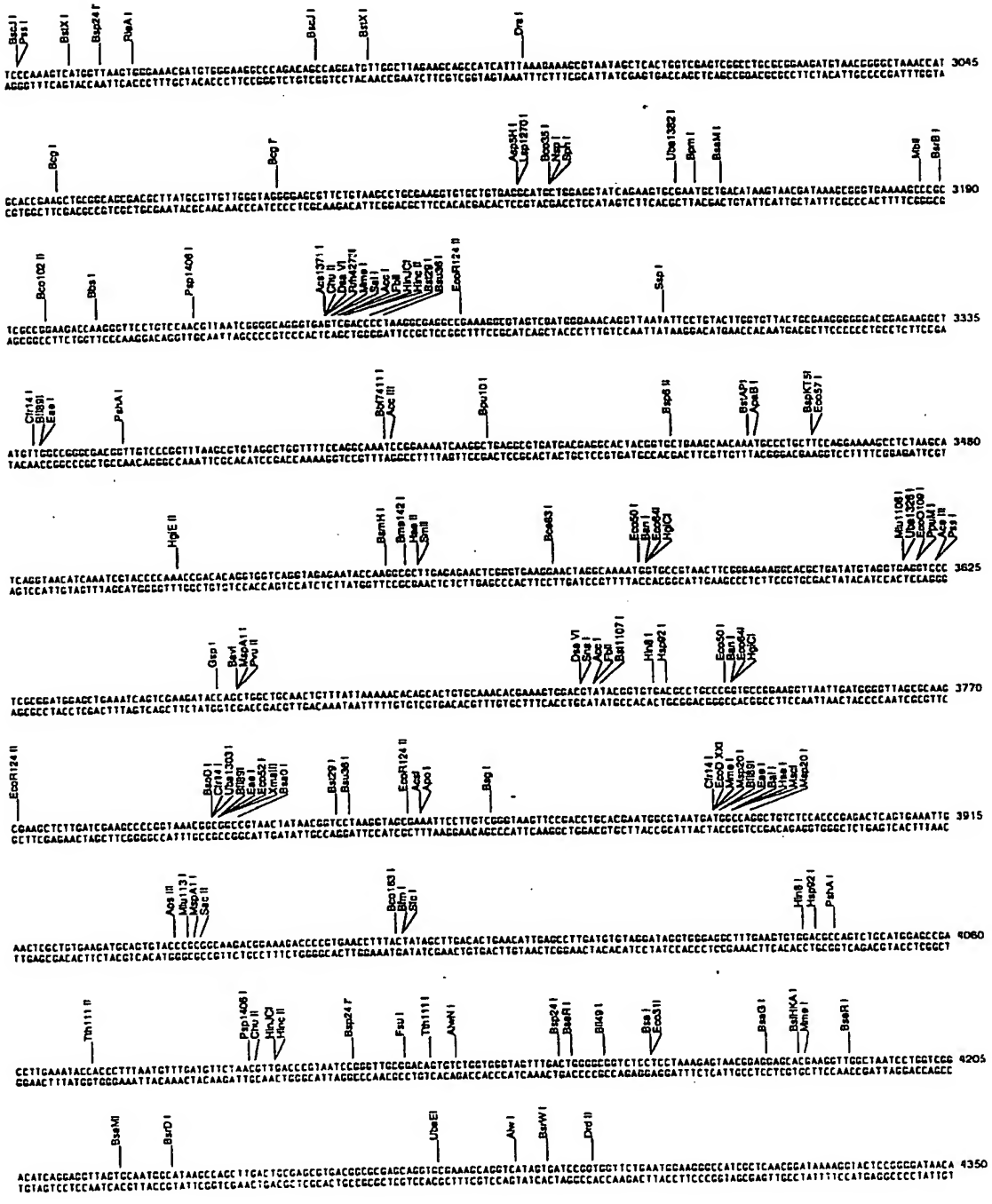


Fig. 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnE* operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less (continued 3).

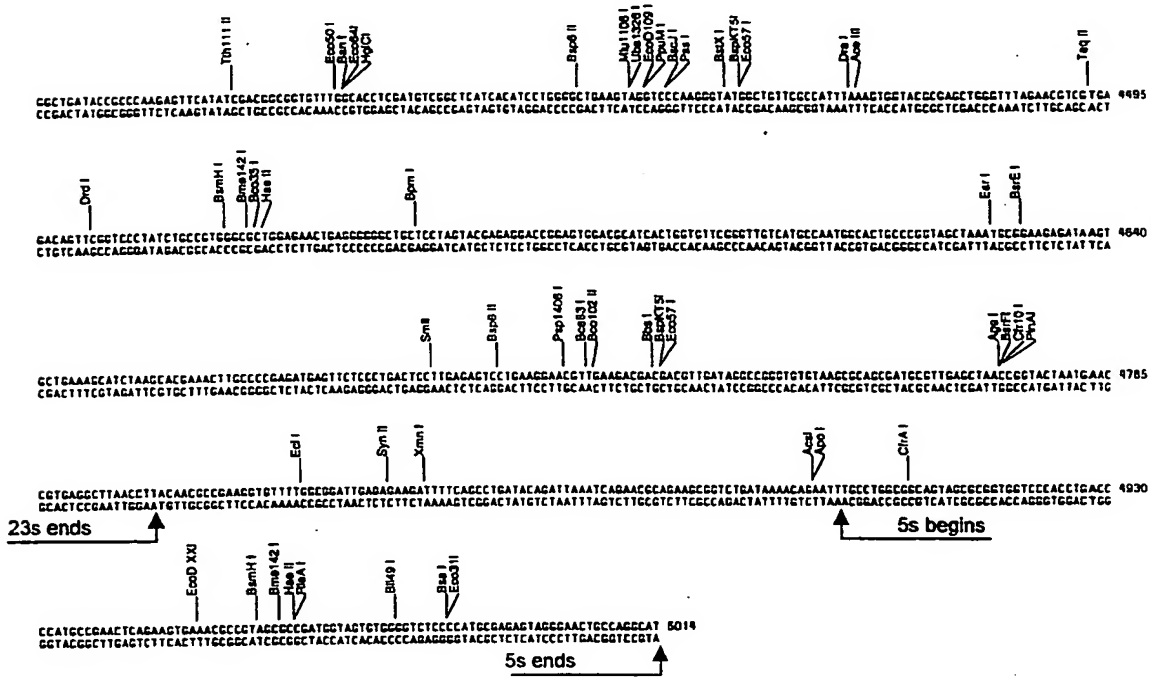


Fig. 7 cont'd

Restriction map of the *rrnE* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less.

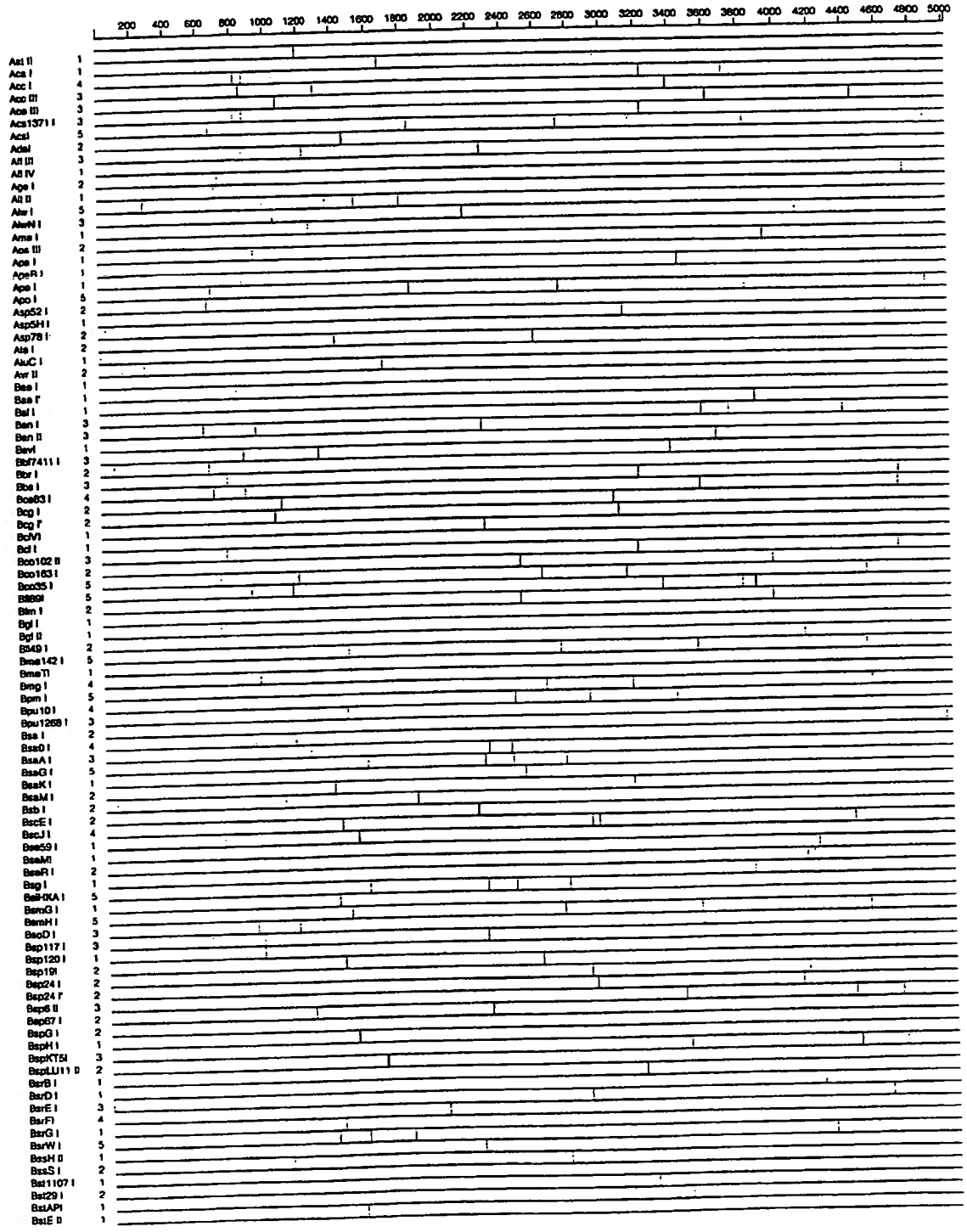


Fig. 7 cont'd

Restriction map of the *rrnE* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less (continued 1).

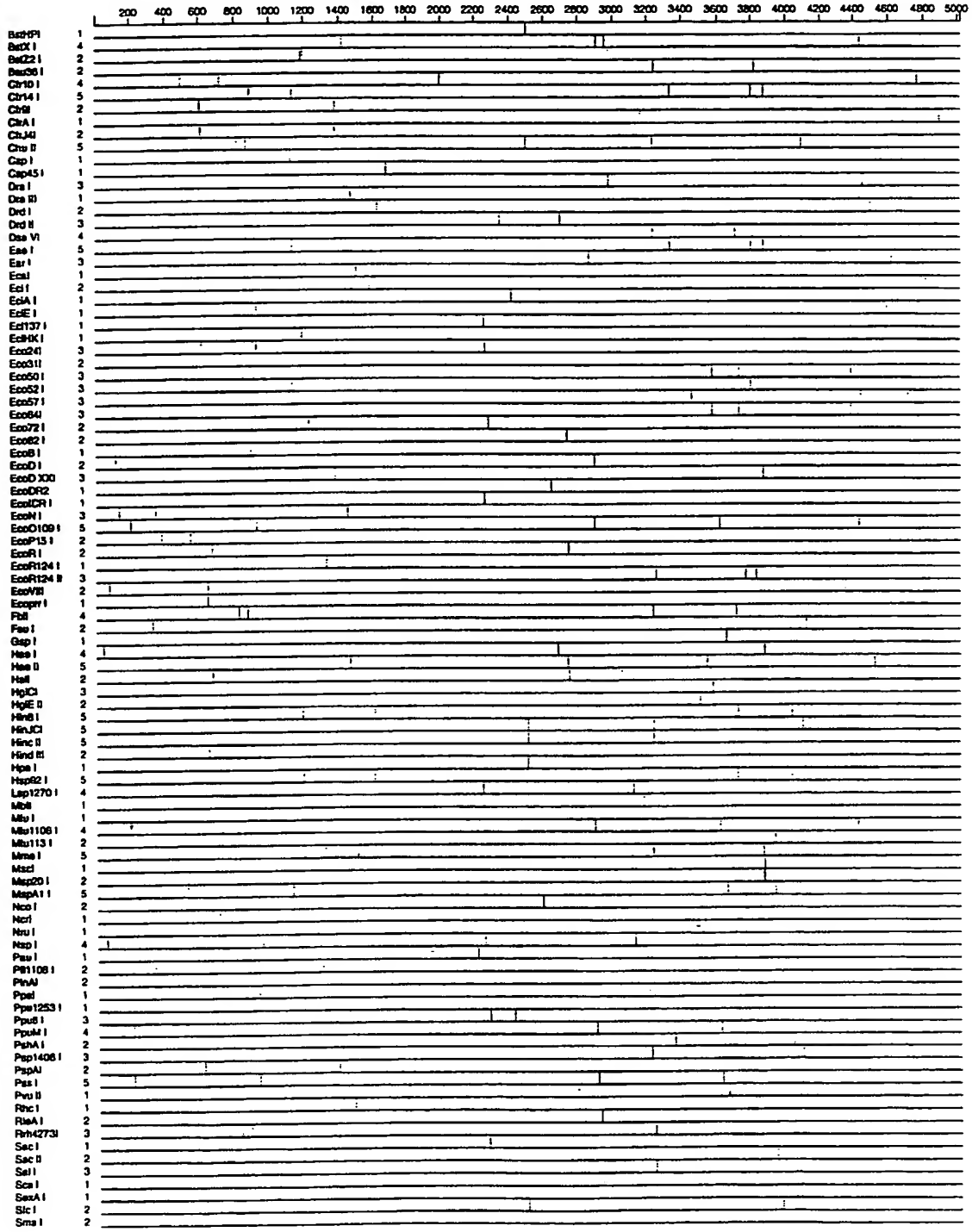


Fig. 7 cont'd

Restriction map of the *rrnE* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less (continued 2).

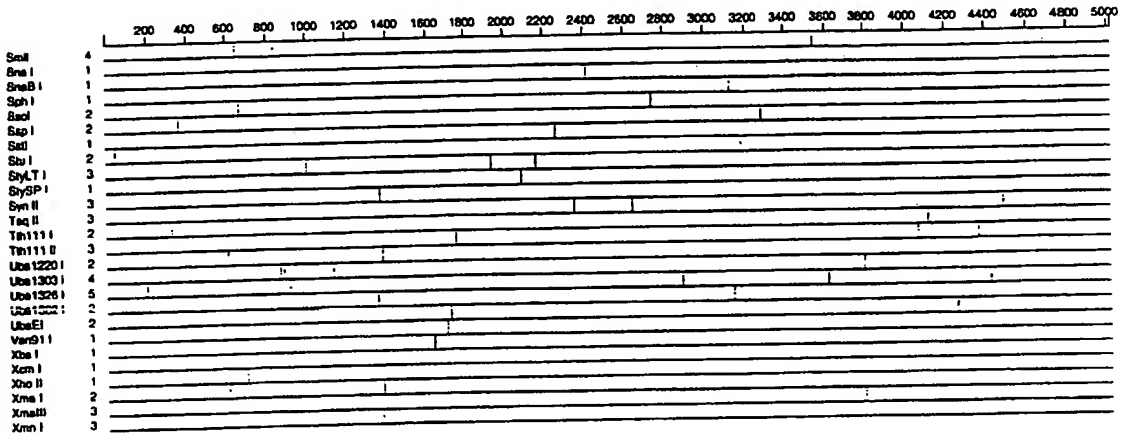


Fig. 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnG* operon (5s-spacer-23s-spacer-16s), with restriction sites noted for enzymes cutting 5 times or less.

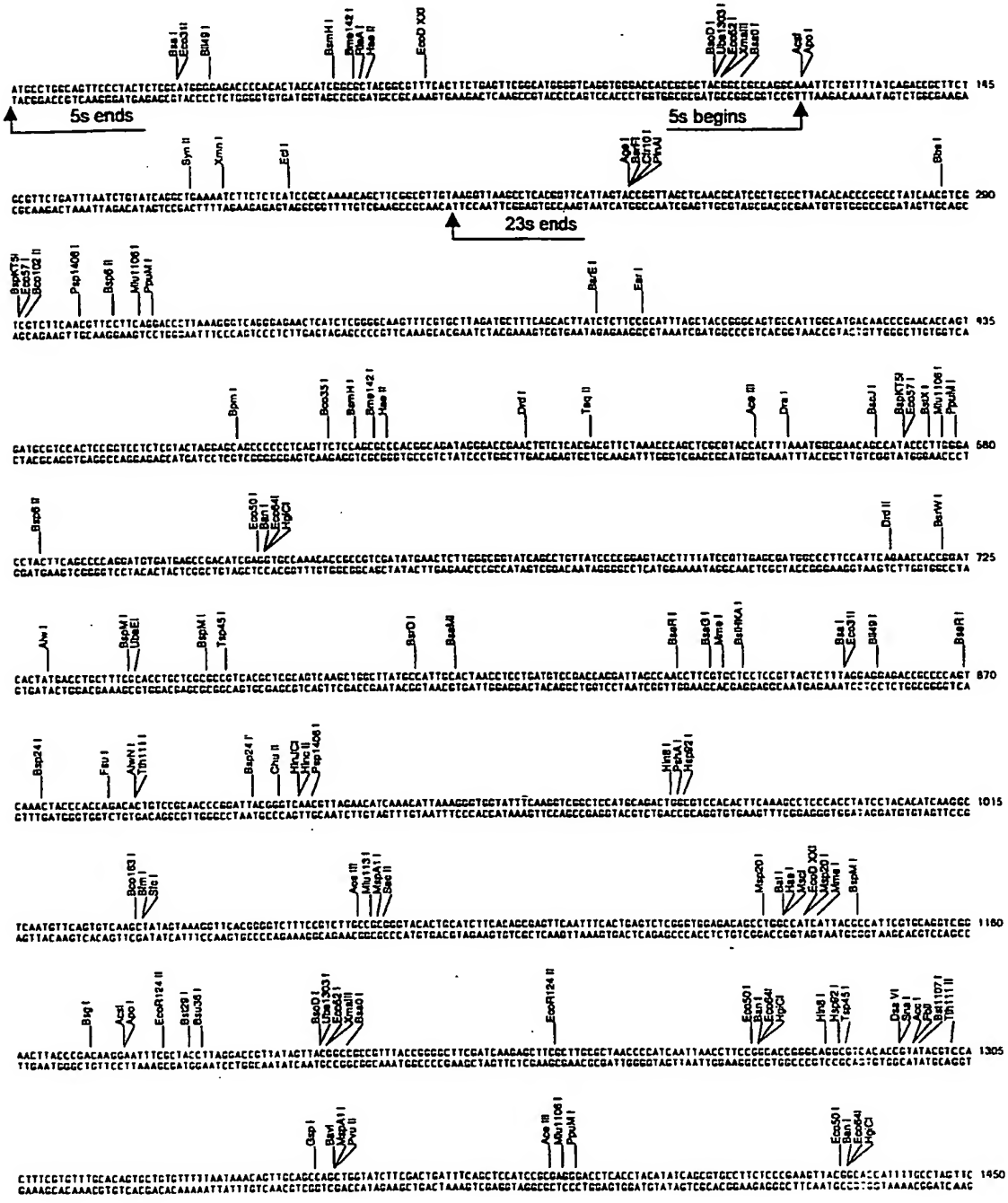


Fig. 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnG* operon (5s-spacer-23s-spacer-16s), with restriction sites noted for enzymes cutting 5 times or less (continued 1).

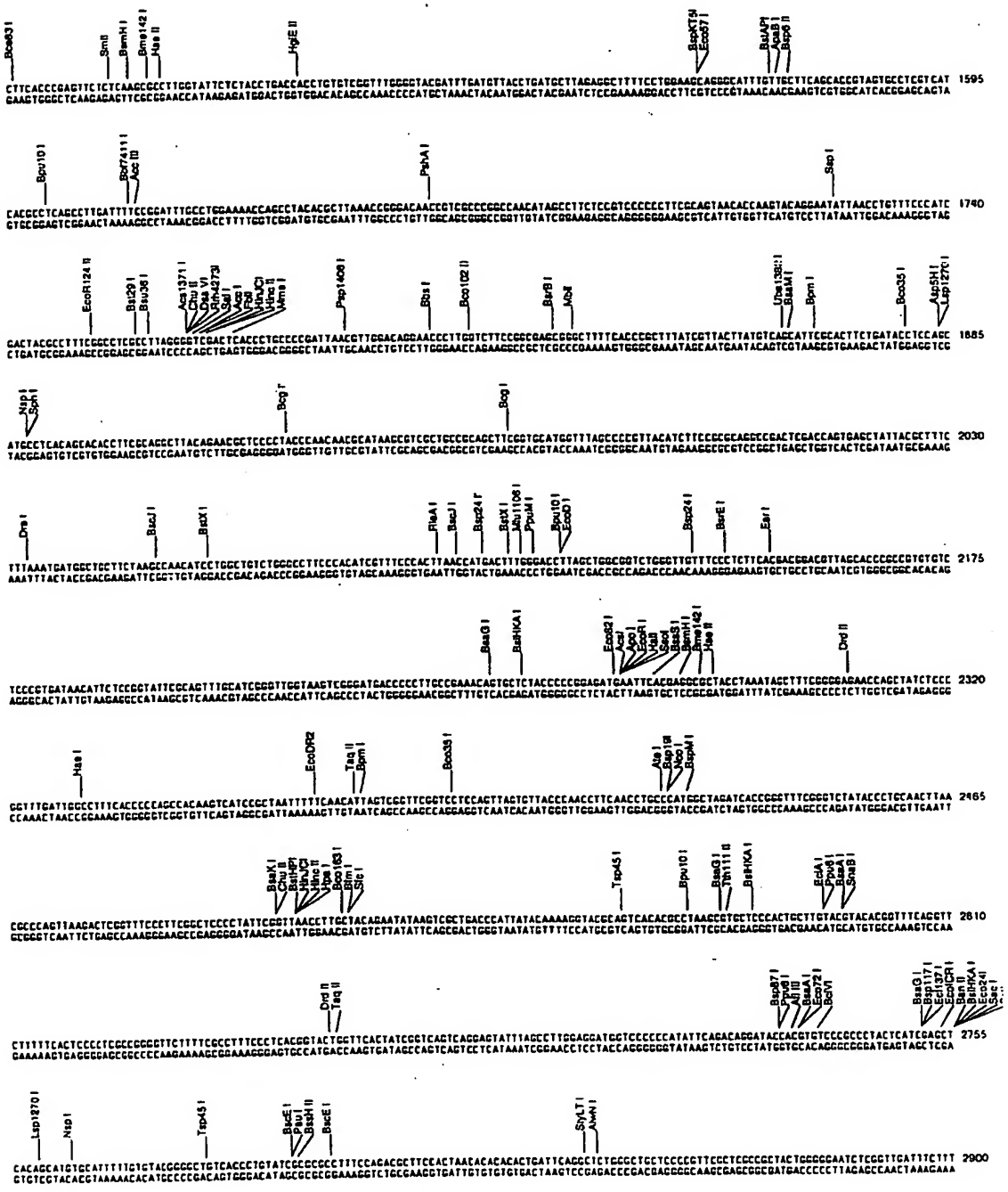


Fig 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnG* operon (5s-spacer-23s-spacer-16s), with restriction sites noted for enzymes cutting 5 times or less (continued 2).

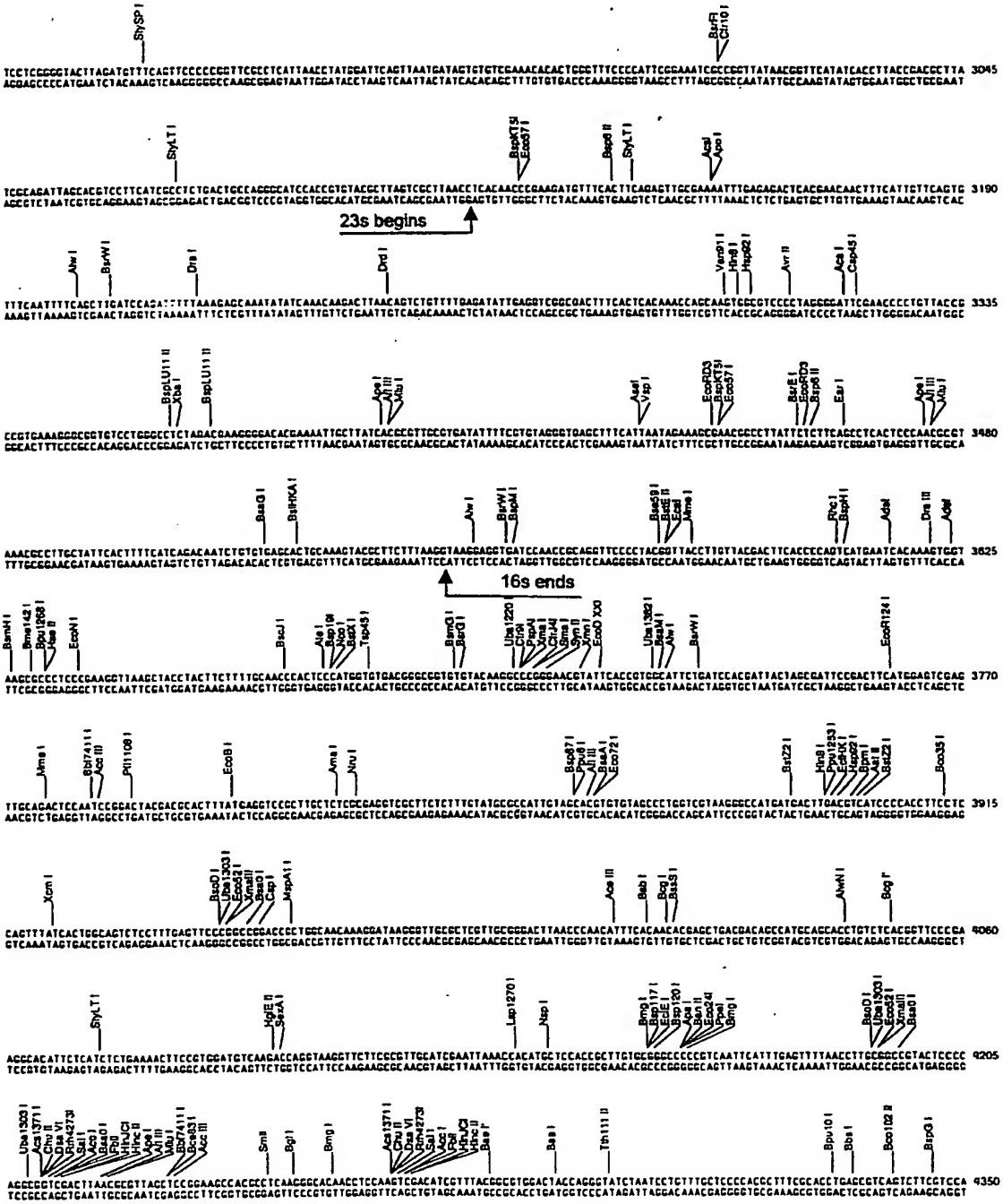


Fig. 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnG* operon (5s-spacer-23s-spacer-16s), with restriction sites noted for enzymes cutting 5 times or less (continued 3).

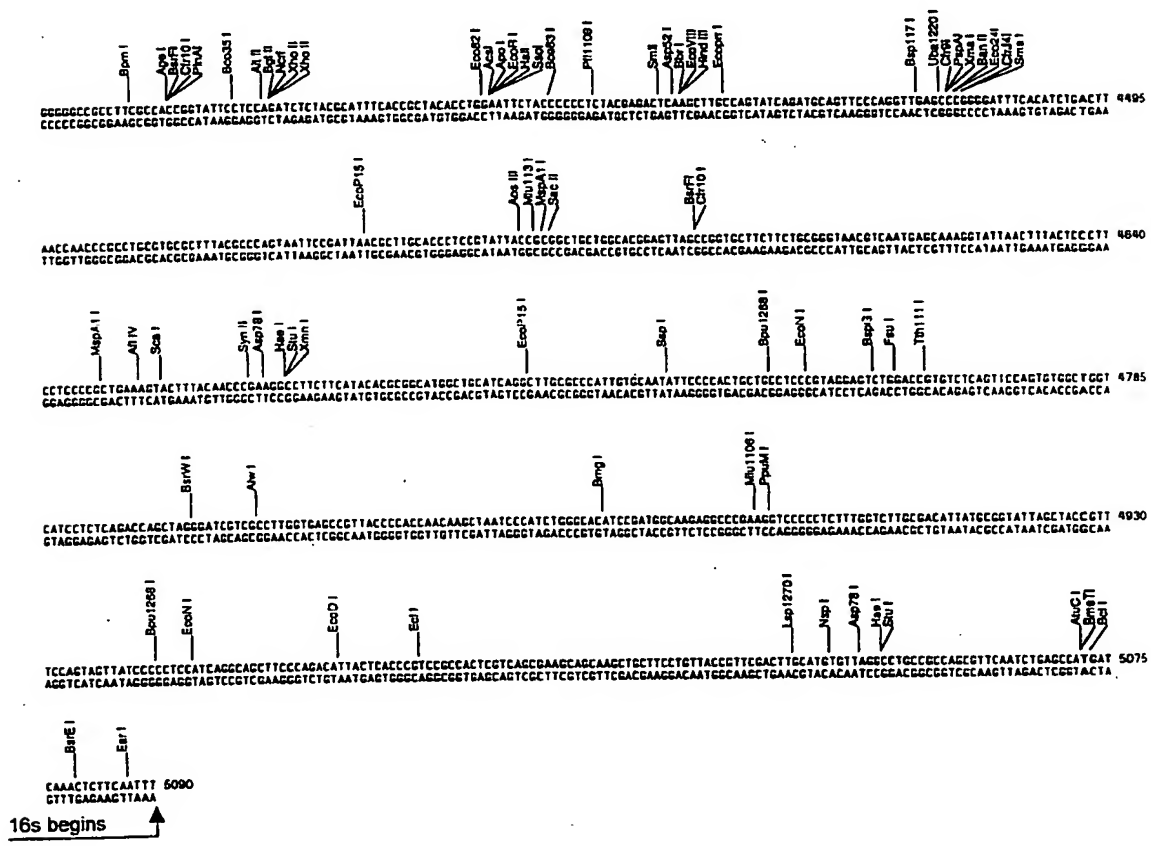


Fig. 7 cont'd

Restriction map of the *rrnG* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less (continued 1).

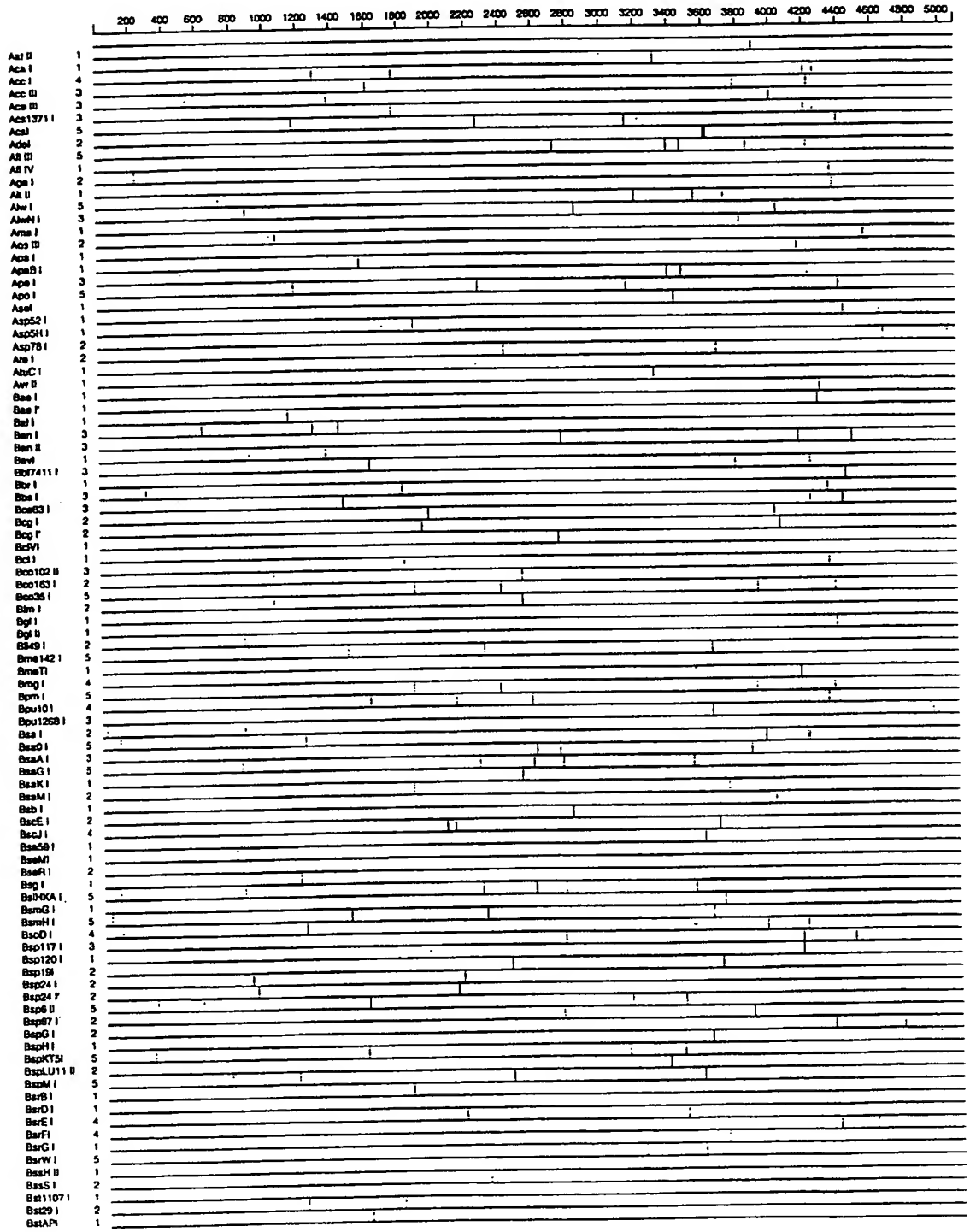


Fig. 7 cont'd

Restriction map of the *rrnG* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less.

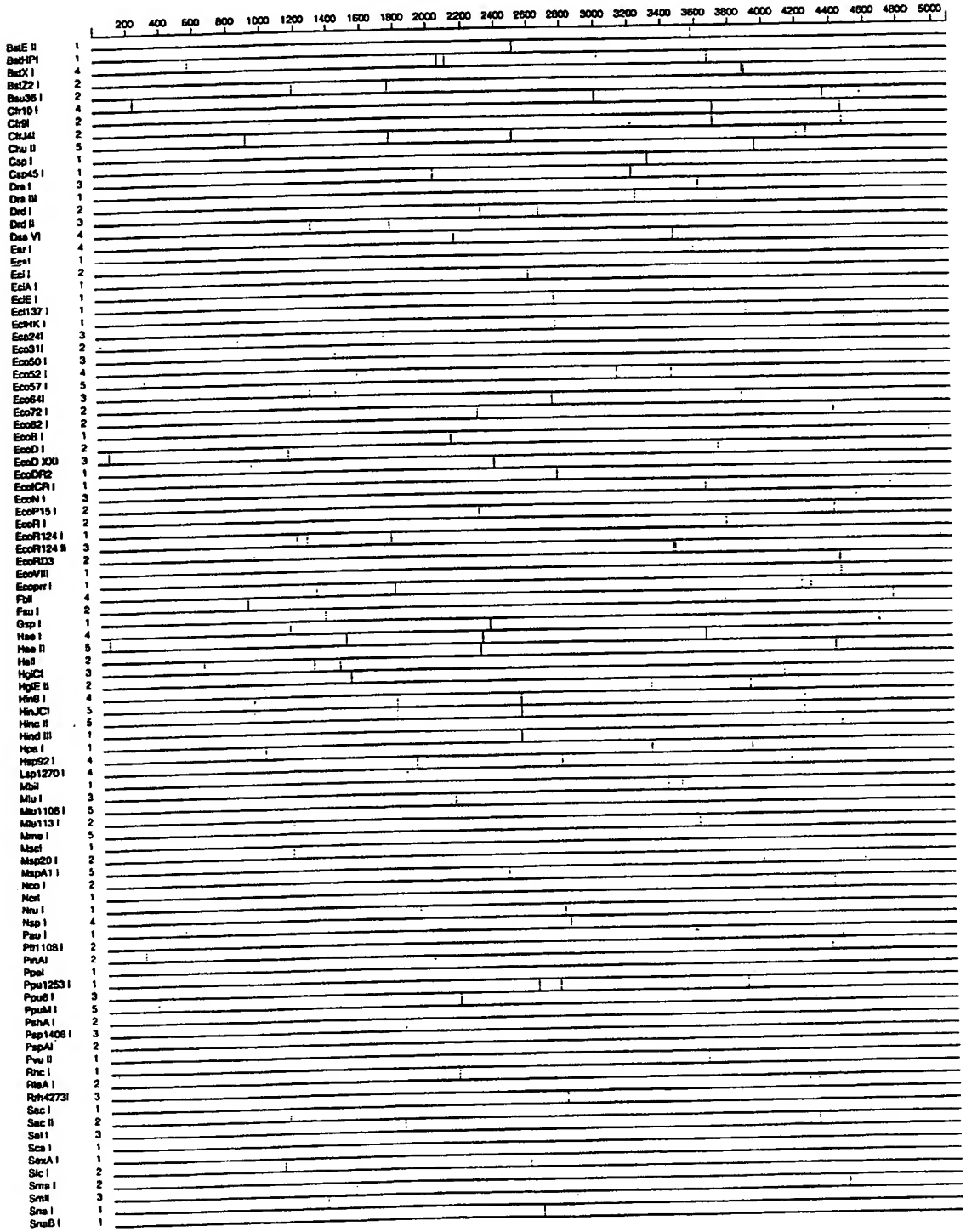


Fig. 7 cont'd

Restriction map of the *rrnG* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less (continued 2).

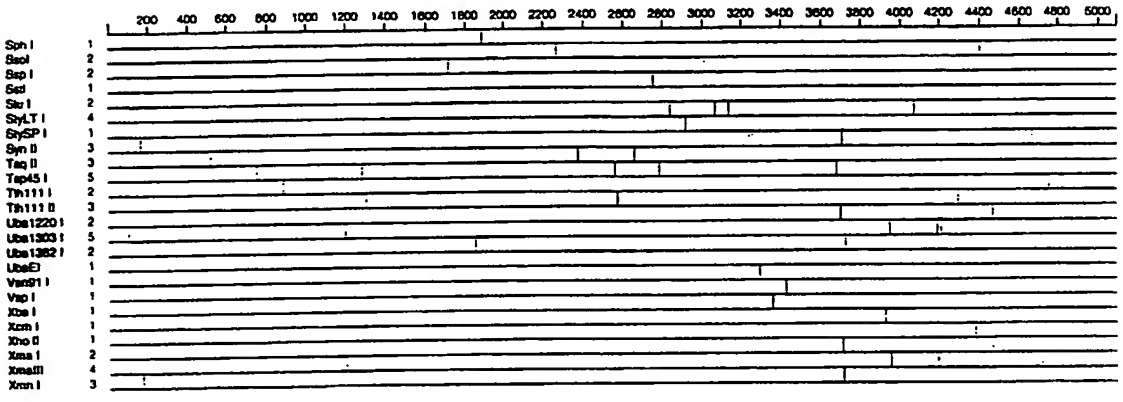


Fig. 7 Cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnH* operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less.

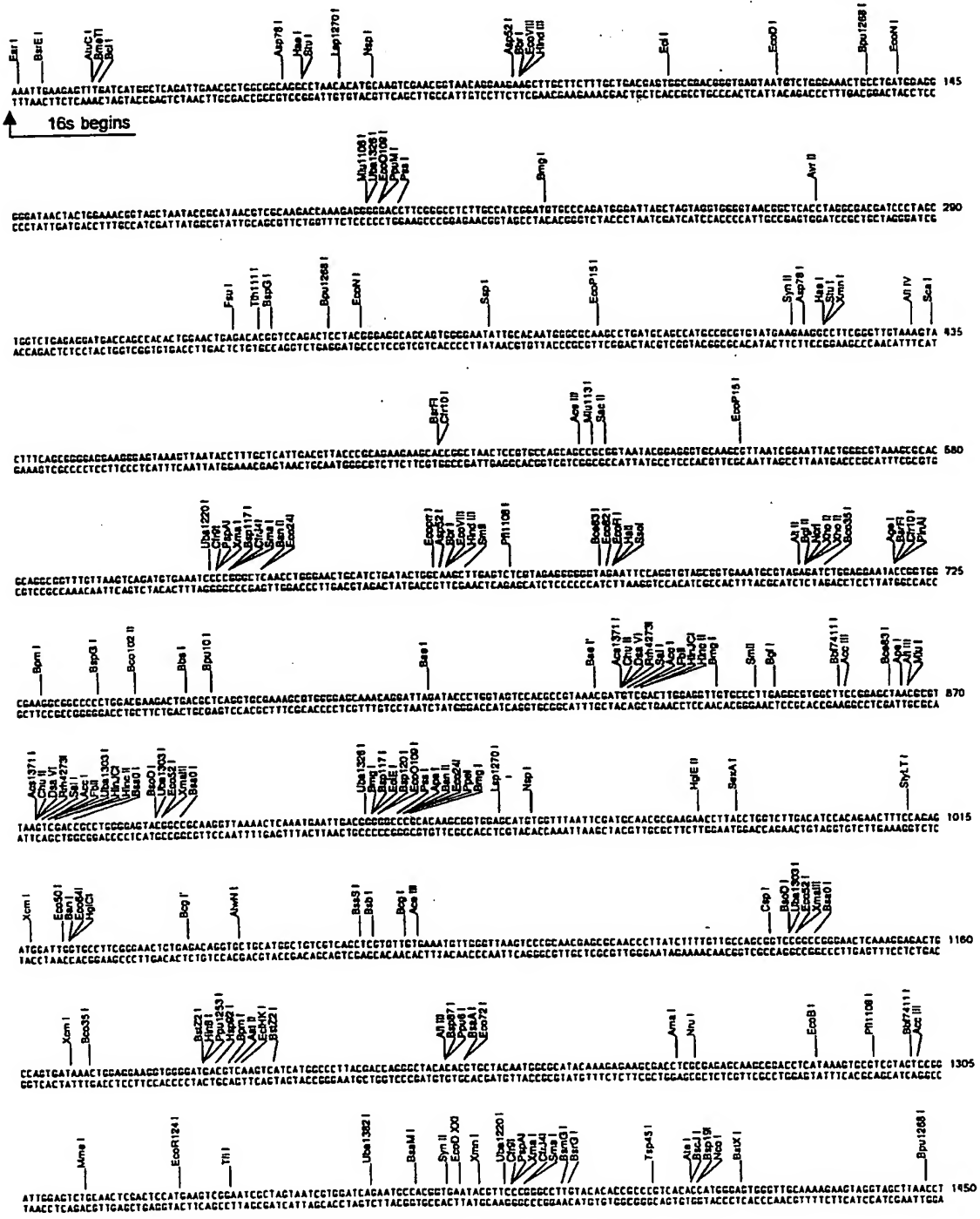


Fig. 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnH* operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less (continued 1).

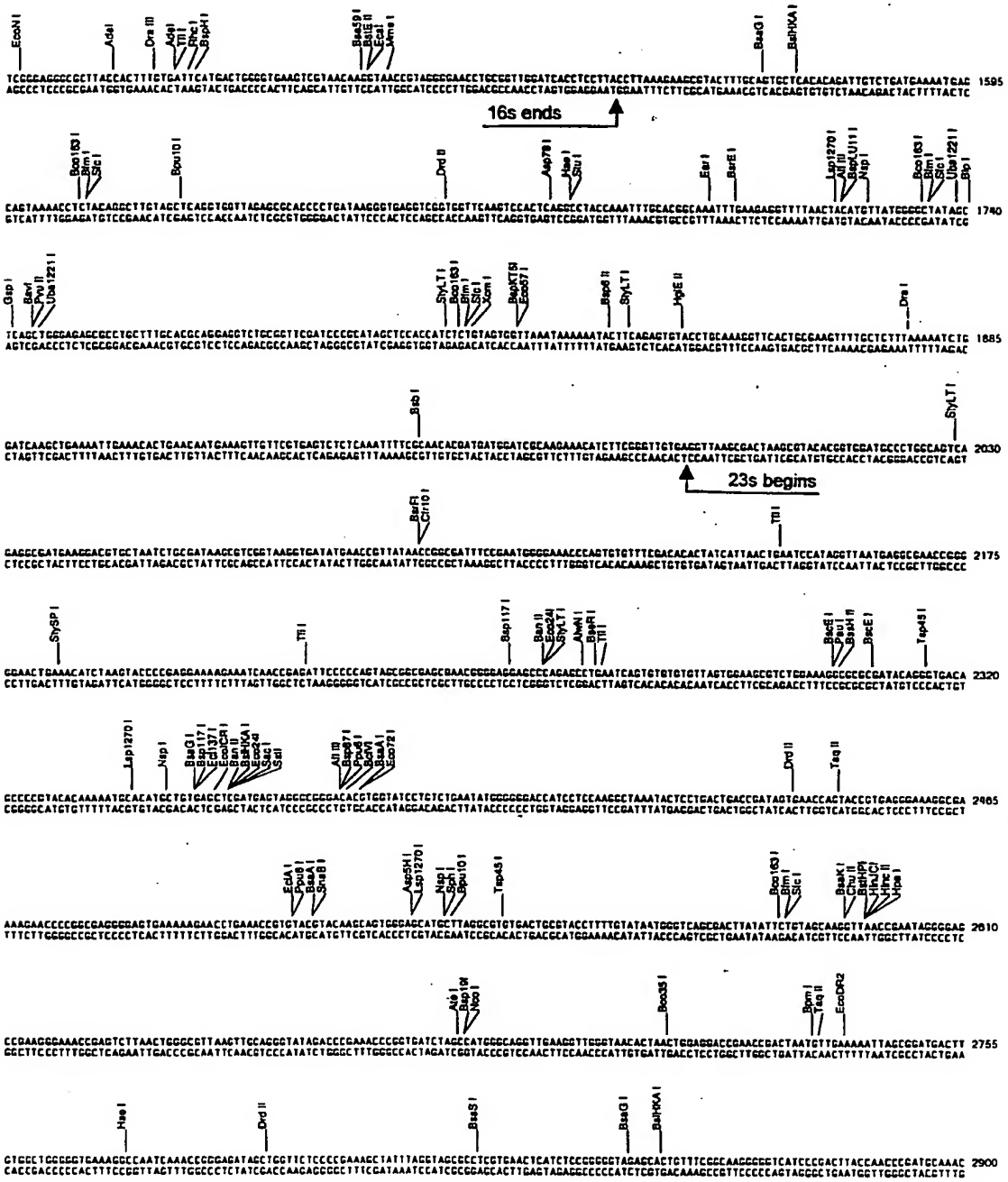


Fig. 7 cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnH* operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less (continued 2).

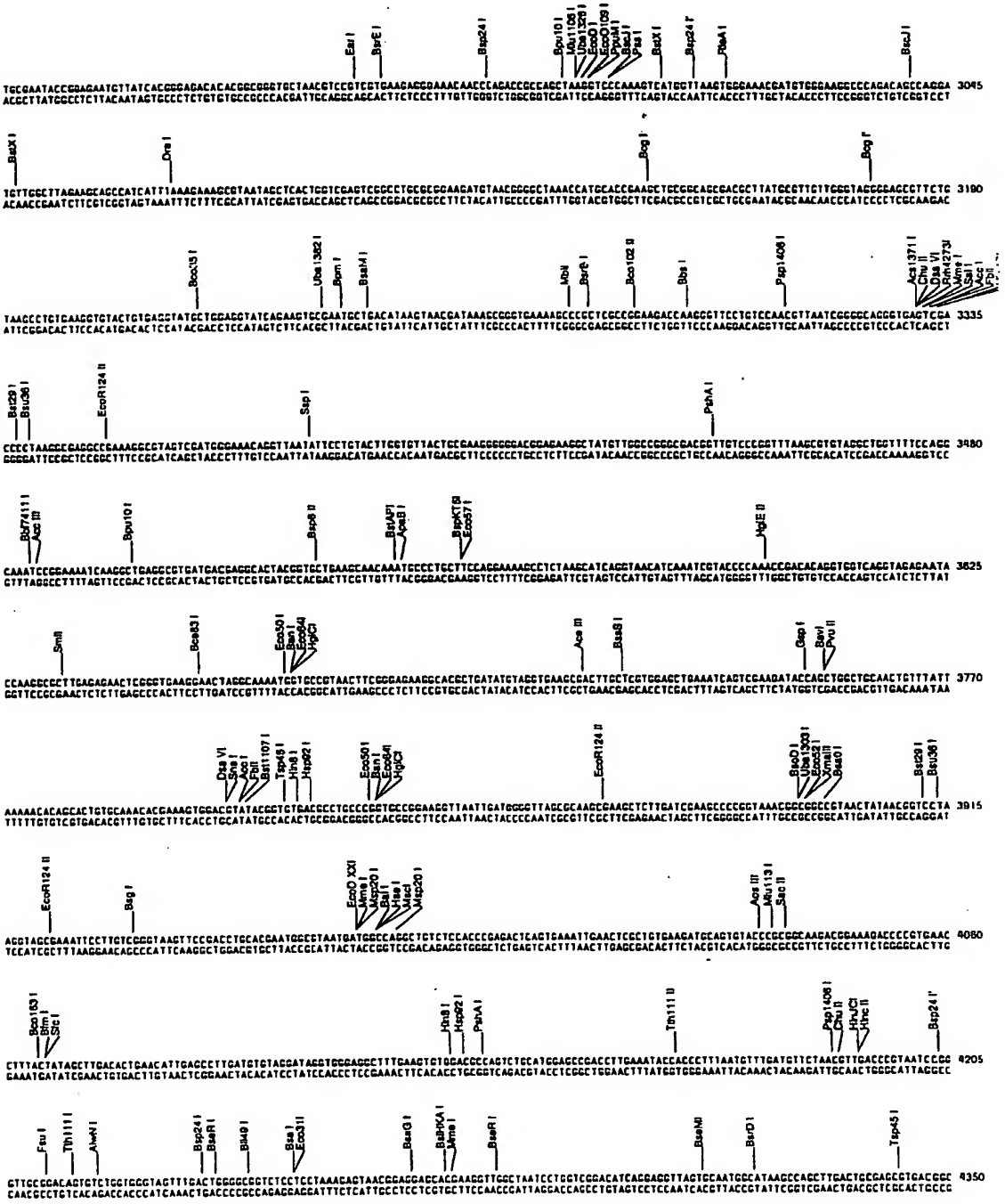


Fig. 7 Cont'd

DNA sequence and restriction map of the *E. coli* MG1655 *rrnH* operon (16s-spacer-23s-spacer-5s), with restriction sites noted for enzymes cutting 5 times or less (continued 3).

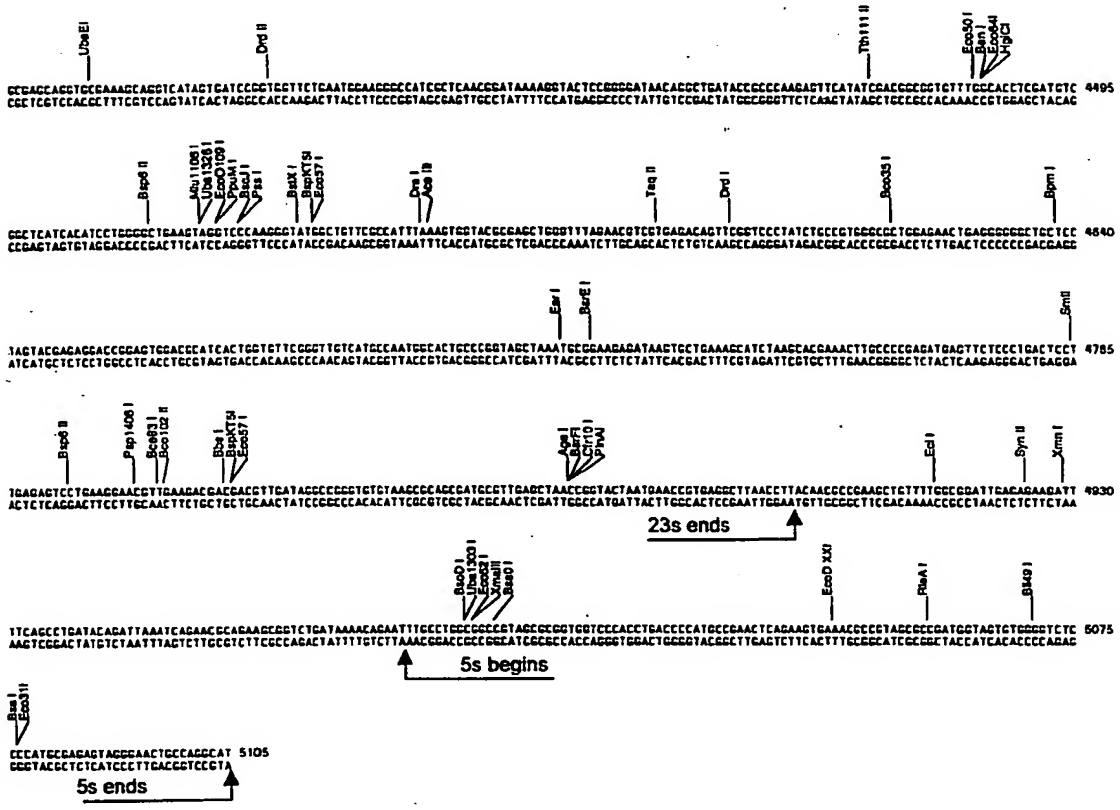


Fig. 7 cont'd

Restriction map of the *rmH* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less.

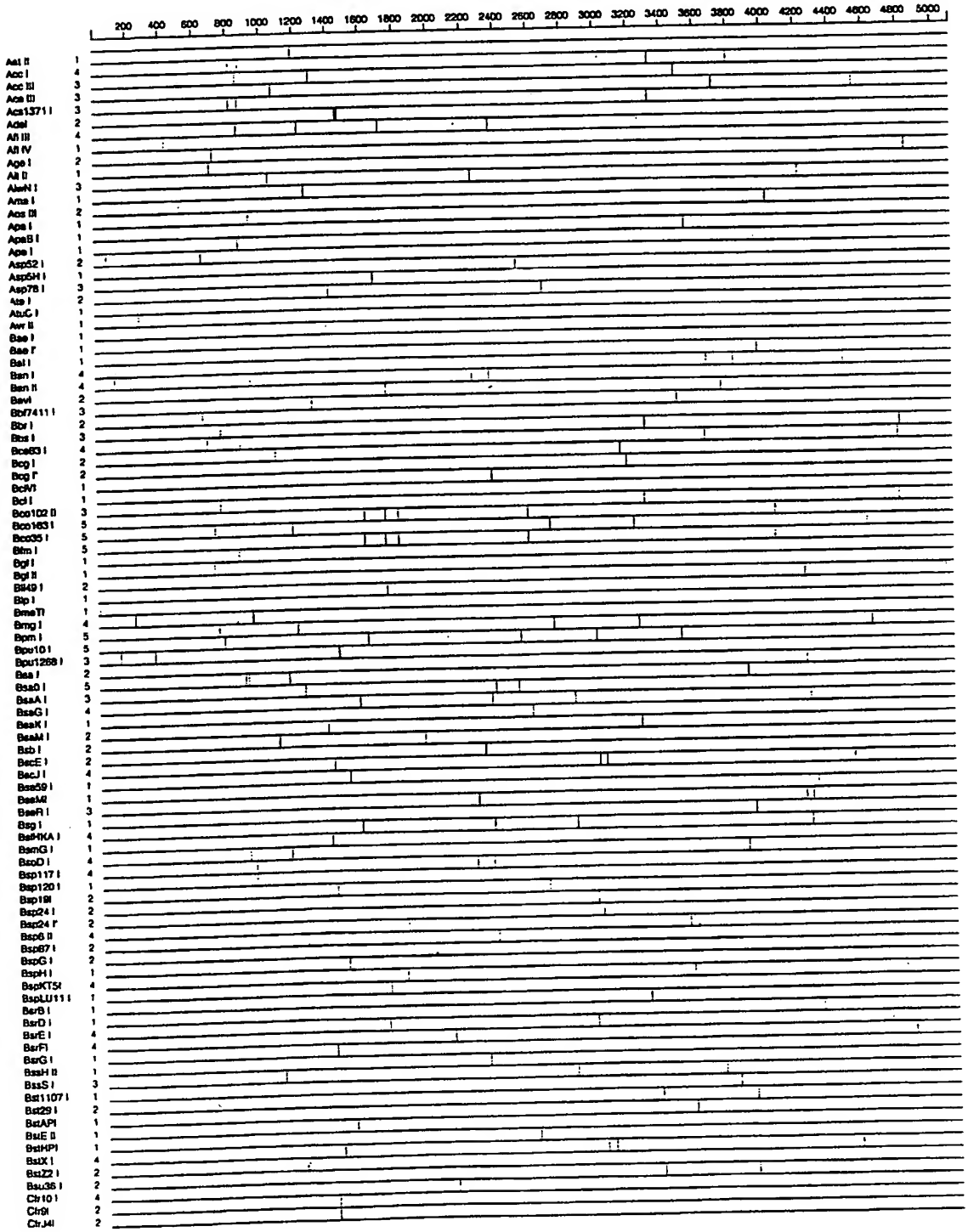


Fig. 7 cont'd

Restriction map of the *rrnH* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less (continued 1).

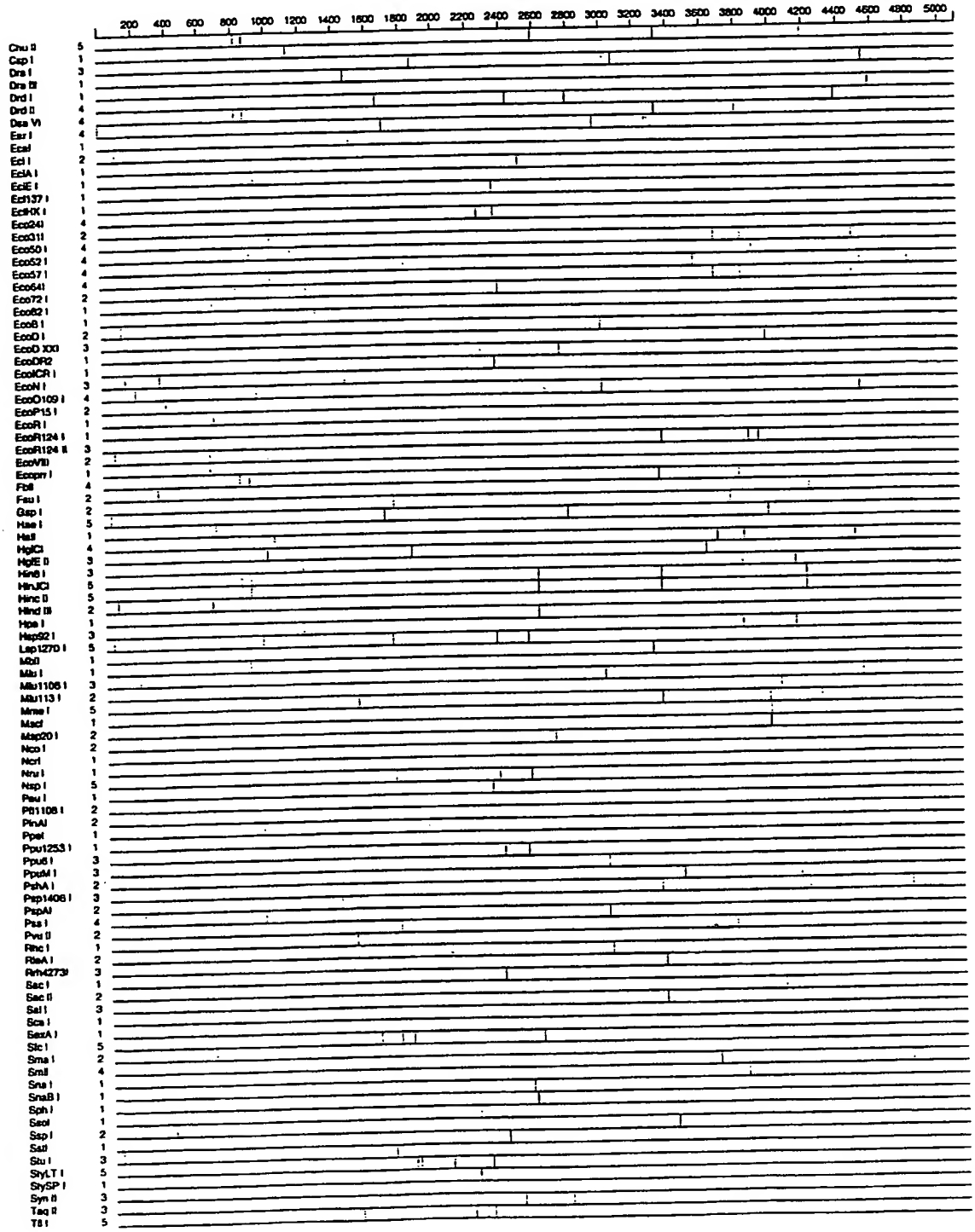
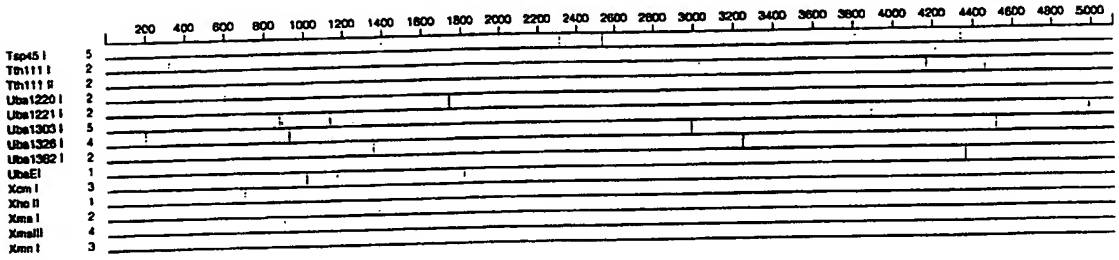


Fig. 7 cont'd

Restriction map of the *rrnH* operon (16s-spacer-23s-spacer-5s) of *E. coli* MG1655, with restriction sites noted for enzymes cutting 5 times or less (continued 2).



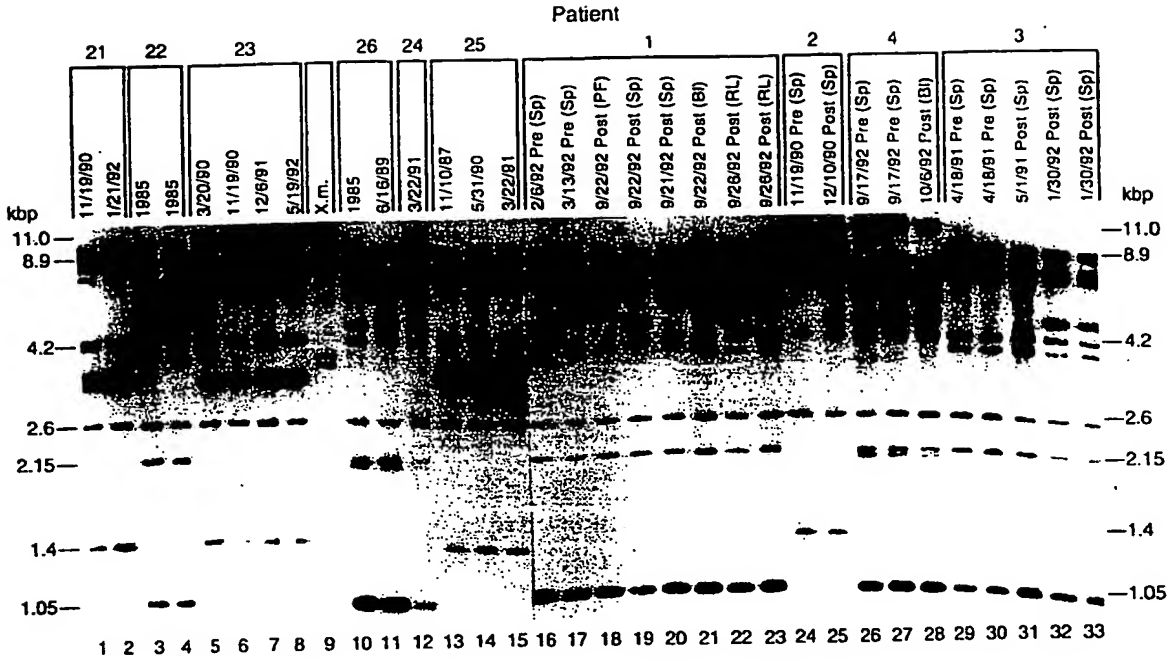

Patient
21
22
23
26
24
25
1
2
4
3
11/19/90
1/21/92
1985
1985
3/20/90
11/19/80
12/6/91
5/19/92
X.m.
1985
9/16/89
3/22/91
11/10/87
5/31/90
3/22/81
2/6/92 Pre (Sp)
3/13/92 Pre (Sp)
9/22/92 Post (PF)
9/22/92 Post (Sp)
9/21/92 Post (Sp)
9/22/92 Post (Bl)
9/28/92 Post (RL)
9/28/92 Post (RL)
11/19/90 Pre (Sp)
12/10/90 Post (Sp)
9/17/92 Pre (Sp)
9/17/92 Pre (Sp)
10/6/92 Post (Bl)
4/18/91 Pre (Sp)
4/18/91 Pre (Sp)
5/1/91 Post (Sp)
1/30/92 Post (Sp)
1/30/92 Post (Sp)
kbp
11.0
8.9
4.2
2.6
2.15
1.4
1.05
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33

FIG. 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/15464

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12Q 1/68; C07H 21/04, 21/02; C12 N 9/12
US CL :435/6, 194; 536/24.32
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 194; 536/24.32

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS: rflps, restriction fragments, bacteria, ma

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 4,717,653 A (WEBSTER, Jr.) 05 January 1988. See entire document.	1-4

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 16 OCTOBER 1998	Date of mailing of the international search report 30 OCT 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer STEPHANIE ZITOMER, PHD Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*